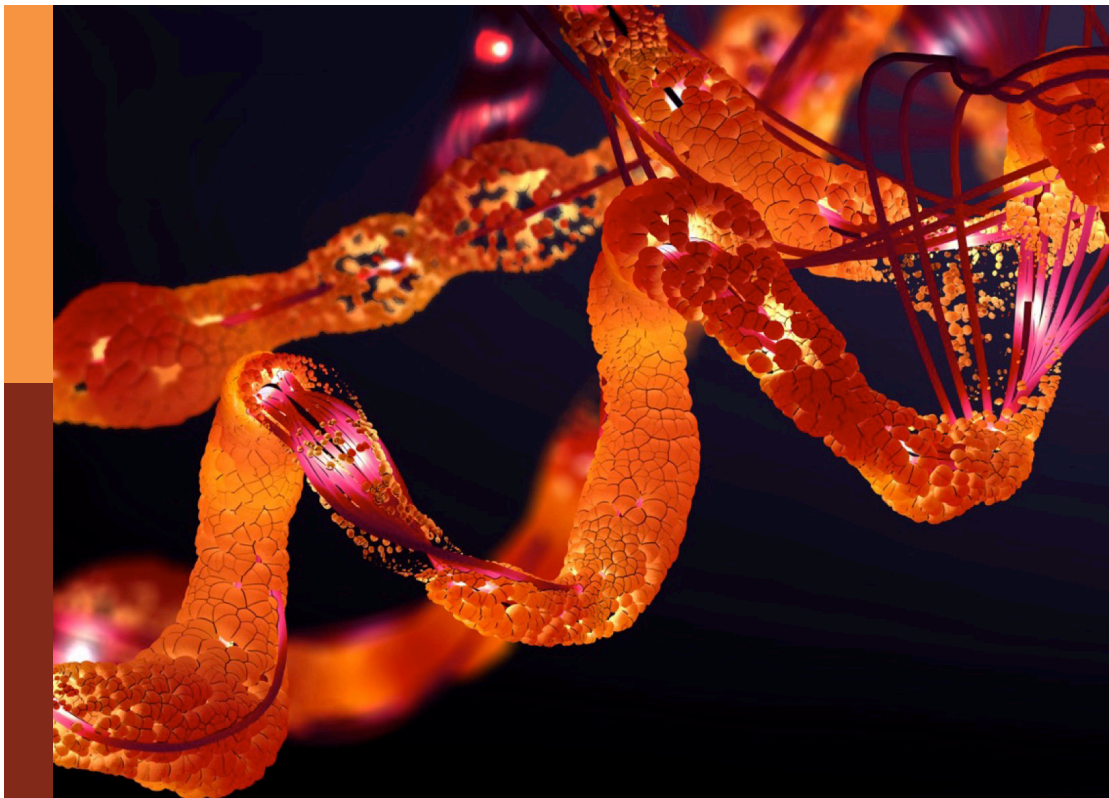
frontiers | Research Topics

Medical knowledge-assisted machine learning technologies in individualized medicine

Edited by
Feng Gao, Xin Gao, Wei Wang and William C. Cho

Published in
Frontiers in Molecular Biosciences
Frontiers in Physiology
Frontiers in Genetics



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 978-2-83252-130-4
DOI 10.3389/978-2-83252-130-4

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

Medical knowledge-assisted machine learning technologies in individualized medicine

Topic editors

Feng Gao — The Sixth Affiliated Hospital of Sun Yat-sen University, China
Xin Gao — King Abdullah University of Science and Technology, Saudi Arabia
Wei Wang — First Affiliated Hospital of Anhui Medical University, China
William C. Cho — QEH, Hong Kong, SAR China

Citation

Gao, F., Gao, X., Wang, W., Cho, W. C., eds. (2023). *Medical knowledge-assisted machine learning technologies in individualized medicine*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83252-130-4

Table of contents

07 Editorial: Medical knowledge-assisted machine learning technologies in individualized medicine
Feng Gao, William C. Cho, Xin Gao and Wei Wang

09 Effects of Inflammatory Response Genes on the Immune Microenvironment in Colorectal Cancer
YaChen Wang, Luping Zhang, Guanghuan Shi, Mingqing Liu, Weidan Zhao, Yingli Zhang, Ying Wang and Nan Zhang

22 Artificial Intelligence Assisting the Early Detection of Active Pulmonary Tuberculosis From Chest X-Rays: A Population-Based Study
Mayidili Nijiati, Jie Ma, Chuling Hu, Abudouresuli Tuersun, Abudoukeyoumujiang Abulizi, Abudoureyimu Kelimu, Dongyu Zhang, Guanbin Li and Xiaoguang Zou

31 DNA Repair–Related Gene Signature in Predicting Prognosis of Colorectal Cancer Patients
Min-Yi Lv, Wei Wang, Min-Er Zhong, Du Cai, Dejun Fan, Cheng-Hang Li, Wei-Bin Kou, Ze-Ping Huang, Xin Duan, Chuling Hu, Qiqi Zhu, Xiaosheng He and Feng Gao

44 Endoscopy-Based Deep Convolutional Neural Network Predicts Response to Neoadjuvant Treatment for Locally Advanced Rectal Cancer
Xijie Chen, Junguo Chen, Xiaosheng He, Liang Xu, Wei Liu, Dezheng Lin, Yuxuan Luo, Yue Feng, Lei Lian, Jiancong Hu and Ping Lan

55 Predicting the Stone-Free Status of Percutaneous Nephrolithotomy With the Machine Learning System: Comparative Analysis With Guy's Stone Score and the S.T.O.N.E Score System
Hong Zhao, Wanling Li, Junsheng Li, Li Li, Hang Wang and Jianming Guo

61 Multi-Size Deep Learning Based Preoperative Computed Tomography Signature for Prognosis Prediction of Colorectal Cancer
Cheng-Hang Li, Du Cai, Min-Er Zhong, Min-Yi Lv, Ze-Ping Huang, Qiqi Zhu, Chuling Hu, Haoning Qi, Xiaojian Wu and Feng Gao

71 S-Thyroid Computer-Aided Diagnosis Ultrasound System of Thyroid Nodules: Correlation Between Transverse and Longitudinal Planes
Keen Yang, Jing Chen, Huaiyu Wu, Hongtian Tian, Xiuqin Ye, Jinfeng Xu, Xunpeng Luo and Fajin Dong

80 Construction of an HLA Classifier for Early Diagnosis, Prognosis, and Recognition of Immunosuppression in Sepsis by Multiple Transcriptome Datasets
Zhen Chen, Rui Chen, Yangpeng Ou, Jianhai Lu, Qianhua Jiang, Genglong Liu, Liping Wang, Yayun Liu, Zhujiang Zhou, Ben Yang and Liuer Zuo

102 Identification and Validation of lncRNA-AC087588.2 in Lung Adenocarcinoma: A Novel Prognostic and Diagnostic Indicator
Xiulin Jiang, Xi Chen, Jishu Guo, Fan Zhou, Jun Pu, Luciano Mutti and Xiaoqun Niu

116 ESRRG, ATP4A, and ATP4B as Diagnostic Biomarkers for Gastric Cancer: A Bioinformatic Analysis Based on Machine Learning
Qiu Chen, Yu Wang, Yongjun Liu and Bin Xi

130 Multi-Model Domain Adaptation for Diabetic Retinopathy Classification
Guanghua Zhang, Bin Sun, Zhaoxia Zhang, Jing Pan, Weihua Yang and Yunfang Liu

140 A Predictive Model for the 10-year Overall Survival Status of Patients With Distant Metastases From Differentiated Thyroid Cancer Using XGBoost Algorithm-A Population-Based Analysis
Shuai Jin, Xing Yang, Quliang Zhong, Xiangmei Liu, Tao Zheng, Lingyan Zhu and Jingyuan Yang

151 *MetastaSite*: Predicting metastasis to different sites using deep learning with gene expression data
Somayah Albaradei, Abdurhman Albaradei, Asim Alsaedi, Mahmut Uludag, Maha A. Thafar, Takashi Gojobori, Magbubah Essack and Xin Gao

167 Deep learning predicts immune checkpoint inhibitor-related pneumonitis from pretreatment computed tomography images
Peixin Tan, Wei Huang, Lingling Wang, Guanhua Deng, Ye Yuan, Shili Qiu, Dong Ni, Shasha Du and Jun Cheng

178 Construction and validation of an angiogenesis-related gene expression signature associated with clinical outcome and tumor immune microenvironment in glioma
Tianhao Hu, Yutao Wang, Xiaoliang Wang, Run Wang, Yifu Song, Li Zhang and Sheng Han

198 Medical terminology-based computing system: a lightweight post-processing solution for out-of-vocabulary multi-word terms
Nadia Saeed and Hammad Naveed

211 Integrating cell cycle score for precise risk stratification in ovarian cancer
Lingying Chen, Haiyan Gu, Lei Zhou, Jingna Wu, Changdong Sun and Yonggui Han

225 Efficient and accurate diagnosis of otomycosis using an ensemble deep-learning model
Chenggang Mao, Aimin Li, Jing Hu, Pengjun Wang, Dan Peng, Juehui Wang and Yi Sun

236 PTBPs: An immunomodulatory-related prognostic biomarker in pan-cancer
Chen Chen, Anquan Shang, Yuting Gao, Jingjuan Huang, Gege Liu, William C. Cho and Dong Li

254 Computer-aided diagnosis of prostate cancer based on deep neural networks from multi-parametric magnetic resonance imaging
Zhenglin Yi, Zhenyu Ou, Jiao Hu, Dongxu Qiu, Chao Quan, Belaydi Othmane, Yongjie Wang and Longxiang Wu

271 Multimodal ultrasound fusion network for differentiating between benign and malignant solid renal tumors
Dongmei Zhu, Junyu Li, Yan Li, Ji Wu, Lin Zhu, Jian Li, Zimo Wang, Jinfeng Xu, Fajin Dong and Jun Cheng

280 Prognostic value of LECT2 and relevance to immune infiltration in hepatocellular carcinoma
Jiangfeng Qin, Weijie Sun, Hui Zhang, Zihao Wu, Jiapei Shen, Wenhai Wang, Yuanyuan Wei, Yanyan Liu, Yufeng Gao and Honghai Xu

296 Automated measurement of endometrial peristalsis in cine transvaginal ultrasound images
Yue Wang, Xiaokun Li, Niya Wei, Yuanxi Liu, Xinting Liu, Ruijie Sun, Chan Huang, Bin Yao and Huifang Wang

305 Resting-state EEG-based convolutional neural network for the diagnosis of depression and its severity
Mengqian Li, Yuan Liu, Yan Liu, Changqin Pu, Ruocheng Yin, Ziqiang Zeng, Libin Deng and Xing Wang

317 Predicting clinical response to everolimus in ER+ breast cancers using machine-learning
Aritro Nath, Patrick A. Cosgrove, Jeffrey T. Chang and Andrea H. Bild

329 Multi-atlas segmentation and quantification of muscle, bone and subcutaneous adipose tissue in the lower leg using peripheral quantitative computed tomography
Sokratis Makrogiannis, Azubuike Okorie, Angelo Di Iorio, Stefania Bandinelli and Luigi Ferrucci

347 A symmetric prior knowledge based deep learning model for intracerebral hemorrhage lesion segmentation
Mayidili Nijiati, Abudouresuli Tuersun, Yue Zhang, Qing Yuan, Ping Gong, Abudoukeyoumujiang Abulizi, Awanisa Tuoheti, Adili Abulaiti and Xiaoguang Zou

360 A machine learning framework for scRNA-seq UMI threshold optimization and accurate classification of cell types
Isaac Bishara, Jinfeng Chen, Jason I. Griffiths, Andrea H. Bild and Aritro Nath

372 Machine learning models for predicting one-year survival in patients with metastatic gastric cancer who experienced upfront radical gastrectomy
Cheng Zhang, Yi Zhang, Ya-Hui Yang, Hui Xu, Xiao-Peng Zhang, Zhi-Jun Wu, Min-Min Xie, Ying Feng, Chong Feng and Tai Ma

383 Comprehensive molecular analyses of an autoimmune-related gene predictive model and immune infiltrations using machine learning methods in moyamoya disease
Shifu Li, Ying Han, Qian Zhang, Dong Tang, Jian Li and Ling Weng

frontiers | Frontiers in Molecular Biosciences

TYPE Editorial
PUBLISHED 24 March 2023
DOI 10.3389/fmolb.2023.1167730

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Cecilia Giulivi,
University of California, Davis, United States

*CORRESPONDENCE
Feng Gao,
gaof57@mail.sysu.edu.cn

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 16 February 2023
ACCEPTED 20 February 2023
PUBLISHED 24 March 2023

CITATION
Gao F, Cho WC, Gao X and Wang W (2023), Editorial: Medical knowledge-assisted machine learning technologies in individualized medicine.
Front. Mol. Biosci. 10:1167730.
doi: 10.3389/fmolb.2023.1167730

COPYRIGHT
© 2023 Gao, Cho, Gao 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.

Editorial: Medical knowledge-assisted machine learning technologies in individualized medicine

Feng Gao[1,2]*, William C. Cho[3], Xin Gao[4,5] and Wei Wang[6]

[1]Department of Colorectal Surgery, Department of General Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2]Shanghai Artificial Intelligence Laboratory, Shanghai, China, [3]Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong SAR, China, [4]Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [5]Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [6]Department of Pathology, First Affiliated Hospital of Anhui Medical University, Hefei, China

KEYWORDS

individualized medicine, machine learning, medical knowledge, computational biology, medical image analysis, biomedical data analysis

Editorial on the Research Topic
Medical knowledge-assisted machine learning technologies in individualized medicine

Machine learning (ML), a powerful tool for mining quantitative disease features in large quantities of medical data, is arguably the most important method for extracting, integrating, and modeling patient-specific and disease-specific factors for individualized medicine. ML can provide clinical decision support for disease diagnosis, prognosis, and treatment in practice. In reality, there is no universal ML model or solution for all medical problems and clinical circumstances (Musolf et al., 2022), since both the fundamental biological mechanisms and the modality of the clinically interested medical data differ from disease to disease. In other words, biological and medical knowledge of the specific domain is essential for the development of ML models for the corresponding medical problems (You et al., 2022). For data scientists with rich modeling experience and limited medical background knowledge, considering even the simplest basic medical knowledge will significantly refine the structure of the ML framework and the process of model training, offering an opportunity for the establishment of ML models with higher performance and better interpretability (Chen et al., 2022). From another aspect, for biomedical researchers of different specialties, effectively incorporating ML models with clinical information and biological knowledge will lead to further insight into the pathogeneses of different diseases and new biomedical findings (Jin et al., 2021; Eloranta and Boman, 2022), which may translate into new drugs or therapeutic strategies for precision medicine. This Research Topic collects some of the contributions from both sides, providing insights into the medical knowledge-assisted ML studies for disease diagnosis, prognosis, and individualized treatment management.

Currently, the ML technique is more and more adopted in the field of genomics analysis. Albaradei et al. presented a pan-cancer model that takes the gene expression profile of the primary site and predicts whether and where the metastasis has occurred. According to

previous studies, metastasis does not occur randomly, noting that the relationship between the metastasis clusters/cells and the specific remote organ is similar to the relationship between the seed and soil (Yuan et al., 2019). Following the "seed and soil" hypothesis, Albaradei et al. identified metastasis-related primary tumor transcriptome features of the common metastasis target organs, such as the lung, liver, bone, and brain. Incorporating the metastasis expression features for common metastasis target organs, the established model of this study perform well in pinpointing the metastasis site, aiding decision-making in clinical practice. In another study, Nath et al. developed an integrative ML model predicting everolimus treatment response based on *in vitro* experimental results and gene expression data of a neoadjuvant everolimus therapy cohort. Hormonal therapy is the mainstay of the treatment for estrogen receptor-positive breast cancer, but many estrogen receptor-positive cases resist endocrine therapy (Hanker et al., 2020). The ML model established by Nath et al. integrated experimentally verified everolimus response signatures with features generated by *in silico* analysis of transcriptome data in the clinical cohort, providing accurate everolimus response prediction for individualized breast cancer treatment.

To advance ML-aided precision medicine, this Research Topic also discusses studies of medical images. Images of computer tomography (CT) are vital for the diagnosis, classification, and risk evaluation of intracerebral hemorrhage (ICH) (Li et al., 2021). Nijiati et al. pointed out that one of the most important characteristics of the ICH CT images is the destruction of the symmetrical structure of the brain tissue. With this prior anatomical knowledge, the authors integrated a transformer-based framework to capture the long-distance symmetric information in their Sym-TransNet model, achieving higher ICH lesion segmentation performance than the existing methods. In addition to ICH, CT scans also play an important role in clinical oncology. Li et al. proposed a multi-size convolutional neural network-based model for colorectal cancer recurrence prediction based on CT images. With the help of the human experts' annotation of CT images, the model crop and magnify the inputted CT images at different magnifications to extract radiological features of multiple levels, turning each CT image into a series of images containing both full-image level information and detailed tumoral level information imitating the CT scan reading behavior of the human experts. The model succeeds in predicting patients with high recurrence risk, providing a reliable tool for individualized prognosis evaluation. Except for CT, ultrasonography also plays a part in clinical oncology. Zhu et al. developed a diagnosis model for the evaluation of the malignant degrees of renal tumors based on ultrasound images. The model extract features from both B-mode ultrasound and contrast-enhanced ultrasound-mode images that are annotated by senior radiologists, reaching an expert-level accuracy in differentiating benign cases from malignant cases.

In conclusion, the integration of medical knowledge into ML technology has been regarded as a powerful computer-aided tool for individualized medicine. Aiming at providing a broad view of the fascinating development in the field, this Research Topic collected articles that gave some insights to make good use of ML technology in a clever way to deal with a variety of clinical or biological problems. We envision this Research Topic of articles will provide insights for researchers interested in medical data science, speed up the development of medical knowledge-aided ML methods, and reinforce the clinical applications of the ML models for precision medicine.

Author contributions

Editorial writing: FG and WC; Review and editing: FG, WC, XG, and WW. All authors contributed to the article and approved the submitted version.

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

Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., et al. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Med. Image Anal.* 79, 102444. doi:10.1016/j.media.2022.102444

Eloranta, S., and Boman, M. (2022). Predictive models for clinical decision making: Deep dives in practical machine learning. *J. Intern Med.* 292 (2), 278–295. doi:10.1111/joim.13483

Hanker, A. B., Sudhan, D. R., and Arteaga, C. L. (2020). Overcoming endocrine resistance in breast cancer. *Cancer Cell* 37 (4), 496–513. doi:10.1016/j.ccell.2020.03.009

Jin, S., Zeng, X., Xia, F., Huang, W., and Liu, X. (2021). Application of deep learning methods in biological networks. *Brief. Bioinform* 22 (2), 1902–1917. doi:10.1093/bib/bbaa043

Li, L., Wei, M., Liu, B., Atchaneeyasakul, K., Zhou, F., Pan, Z., et al. (2021). Deep learning for hemorrhagic lesion detection and segmentation on brain CT images. *IEEE J. Biomed. Health Inf.* 25 (5), 1646–1659. doi:10.1109/JBHI.2020.3028243

Musolf, A. M., Holzinger, E. R., Malley, J. D., and Bailey-Wilson, J. E. (2022). What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. *Hum. Genet.* 141 (9), 1515–1528. doi:10.1007/s00439-021-02402-z

You, Y., Lai, X., Pan, Y., Zheng, H., Vera, J., Liu, S., et al. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct. Target Ther.* 7 (1), 156. doi:10.1038/s41392-022-00994-0

Yuan, L., Guo, F., Wang, L., and Zou, Q. (2019). Prediction of tumor metastasis from sequencing data in the era of genome sequencing. *Brief. Funct. Genomics* 18 (6), 412–418. doi:10.1093/bfgp/elz010

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fgene.2022.886949

Effects of Inflammatory Response Genes on the Immune Microenvironment in Colorectal Cancer

YaChen Wang†, Luping Zhang†, Guanghuan Shi, Mingqing Liu, Weidan Zhao, Yingli Zhang, Ying Wang and Nan Zhang**

Department of Gastroenterology, The First Hospital of Jilin University, Changchun, China

OPEN ACCESS

Edited by:
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

Reviewed by:
Zhicheng Zhang,
ByteDance, China
Xiaokun Liang,
Shenzhen Institutes of Advanced Technology (CAS), China

***Correspondence:**
Ying Wang
wy01@jlu.edu.cn
Nan Zhang
zhangnan@jlu.edu.cn

†*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *01 March 2022*
Accepted: *14 March 2022*
Published: *08 April 2022*

Citation:
Wang Y, Zhang L, Shi G, Liu M, Zhao W, Zhang Y, Wang Y and Zhang N (2022) Effects of Inflammatory Response Genes on the Immune Microenvironment in Colorectal Cancer. Front. Genet. 13:886949. doi: 10.3389/fgene.2022.886949

Background: The close relationship between colorectal cancer and inflammation has been widely reported. However, the relationship between colorectal cancer and inflammation at the genetic level is not fully understood.

Method: From a genetic perspective, this study explored the relationship between inflammation-related genes and the immune microenvironment in colorectal cancer. We identified prognostic genes, namely *CX3CL1*, *CCL22*, *SERPINE1*, *LTB4R*, *XCL1*, GAL, *TIMP1*, *ADIPOQ*, and *CRH*, by using univariate and multivariate regression analyses. A risk scoring model for inflammatory response was established, and patients in The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database were divided into two groups: high risk group and low risk group.

Results: The analysis showed that the prognosis of the two groups was significantly different, and the low-risk group had a higher survival rate and longer survival time. Pathways related to apoptosis, inflammatory response, and hypoxia were significantly enriched as shown *via* Gene Set Enrichment Analysis (GSEA). Activated dendritic cell infiltration was found in both the TCGA and GEO databases, and the *CCL21* gene played a significant role in the process of activated dendritic cell infiltration. *CCL21* gene was also positively correlated with inflammatory response, and the gene expression and risk score were significantly different between the two groups.

Conclusion: In summary, inflammatory response has a direct impact on patients with colorectal cancer in the prognosis and immune infiltration and further research studies on the inflammatory response can help in advancing the development of immunotherapy for colorectal cancer.

Keywords: inflammatory response, tumor microenvironment, colorectal cancer, immune cell infiltration, risk score

Abbreviations: AUC, area under the ROC curve; CRC, colorectal cancer; GEO, gene expression omnibus; GSEA, gene set enrichment analysis; NK, natural killer; PPI, protein-protein interaction; TCGA: the cancer genome atlas.

INTRODUCTION

Colorectal cancer is the third most common cancer in the world (Stoffel and Yurgelun, 2016). It is also one of the main causes of cancer deaths in both men and women globally (Drewes et al., 2016). The possible link between inflammation and tumors was first revealed in the 19th century by Rudolf Virchow (Hooper et al., 2012). Epidemiological and clinical studies have also shown that patients who have Crohn's disease and ulcerative colitis, the two major types of inflammatory bowel disease, are at increased risk of colorectal cancer (CRC) (Rizzo et al., 2011). The gut is particularly rich in human microbes, and bacteria disrupt the homeostasis by activating immune signaling pathways, leading to an inflammatory environment (Brennan and Garrett, 2016). Tumor microenvironment plays an important role in the growth and development of tumors (Esteva et al., 2019; Strasser and Birnleitner, 2019). Inflammation can cause the aggregation and activation of immune cells, and the activated immune cells promote the proliferation of tumor cells by secreting pro-inflammatory cytokines and chemokines (Ferrari et al., 2019). Therefore, the relationship between inflammatory response and the immune microenvironment of colorectal cancer has attracted much attention. In this study, we screened out the inflammatory response genes related to the prognosis of intestinal tumors and constructed an inflammatory response risk score model. Enrichment of the model-related pathways was examined. Furthermore, we screened out the genes which may influence immune cell infiltration using the inflammatory response risk score model. The purpose of this study was to examine the relationship between inflammatory response and intestinal tumor immune microenvironment at the genetic level.

MATERIALS AND METHODS

Data Sources

RNA sequences and clinical data of relevance in this study were obtained from the TCGA and GEO (GSE39582) databases. GSE39582 contains the largest sample size of CRC patients with the most complete clinical information.

Construction of the PPI Network

We used a String database to construct the PPI network of genes related to inflammatory response.

Construction and Grouping of Inflammatory Response Models

We downloaded the genes related to inflammatory response, and then screened for the genes associated with the prognosis of colorectal cancer. The expression level of each gene in the TCGA and GEO databases was multiplied by the expression coefficient, followed by calculation of the risk score for each patient. For further analyses, the patients in the two databases were divided into high and low risk groups according to the median value of the risk score obtained from the TCGA database.

Survival Analysis

There were 445 patients in the TCGA database, and the follow-up time was 12 years. There were 562 patients in the GEO database who were followed for 16 years. The "survival" and "SurvMiner" packages from R (4.0.3) language were used to analyze the prognosis of these patients. Kaplan-Meier method was used to draw the survival curves, and the log-rank test was used to test the statistical significance. $p < 0.05$ was considered significant.

ROC Curve Analysis

The 1-, 3-, and 5-years survival rates of patients in two databases were analyzed. The "survival," "SurvMiner," and "timeROC" packages from R (4.0.3) language were used to analyze and calculate the area under the ROC curve (AUC). If the area under the curve of 1-, 3-, and 5-years survival rates gradually increases and exceeds 0.5, it indicates that the model has a high accuracy for predicting the survival of patients. The survival of the two groups was represented by a risk column and risk curve.

Heat Map

The heatmap representing gene expression in this experiment was drawn by the "PheatMap" package of R (4.0.3) language.

Cox Regression Analysis

Survival kit of R (4.0.3) language was used to analyze the inflammatory response genes that were significantly correlated with prognosis, and age, sex, T, N, M, and risk score were used for single-factor and multi-factor prognostic analyses.

Gene Set Enrichment Analysis

We downloaded and extracted the genes associated with inflammation from the 1 Gene Set Enrichment Analysis (GSEA) website. Gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were considered to have statistical significance.

Analysis of the Correlation Between Genes and Inflammatory Response

We downloaded the genes involved in regulating immune cells and screened for the ones that play a crucial role in inflammatory response. The "ggplot2," "GGPUBR," and "ggExtra" packages of R (4.0.3) language were then used to analyze the correlation between these genes and inflammatory response, as well as their expression between the high and low risk groups.

RESULTS

Extraction and Screening of Inflammatory Response-Related Genes

We downloaded the Ontology gene set from the Gene Set Enrichment Analysis (GSEA) website and extracted the inflammatory response-related genes from the data set. We used the STRING protein-protein interaction (PPI) to establish the relationship between the proteins of the inflammatory response-related genes (**Figure 1A**). Because of the large number of genes related to the inflammatory response,

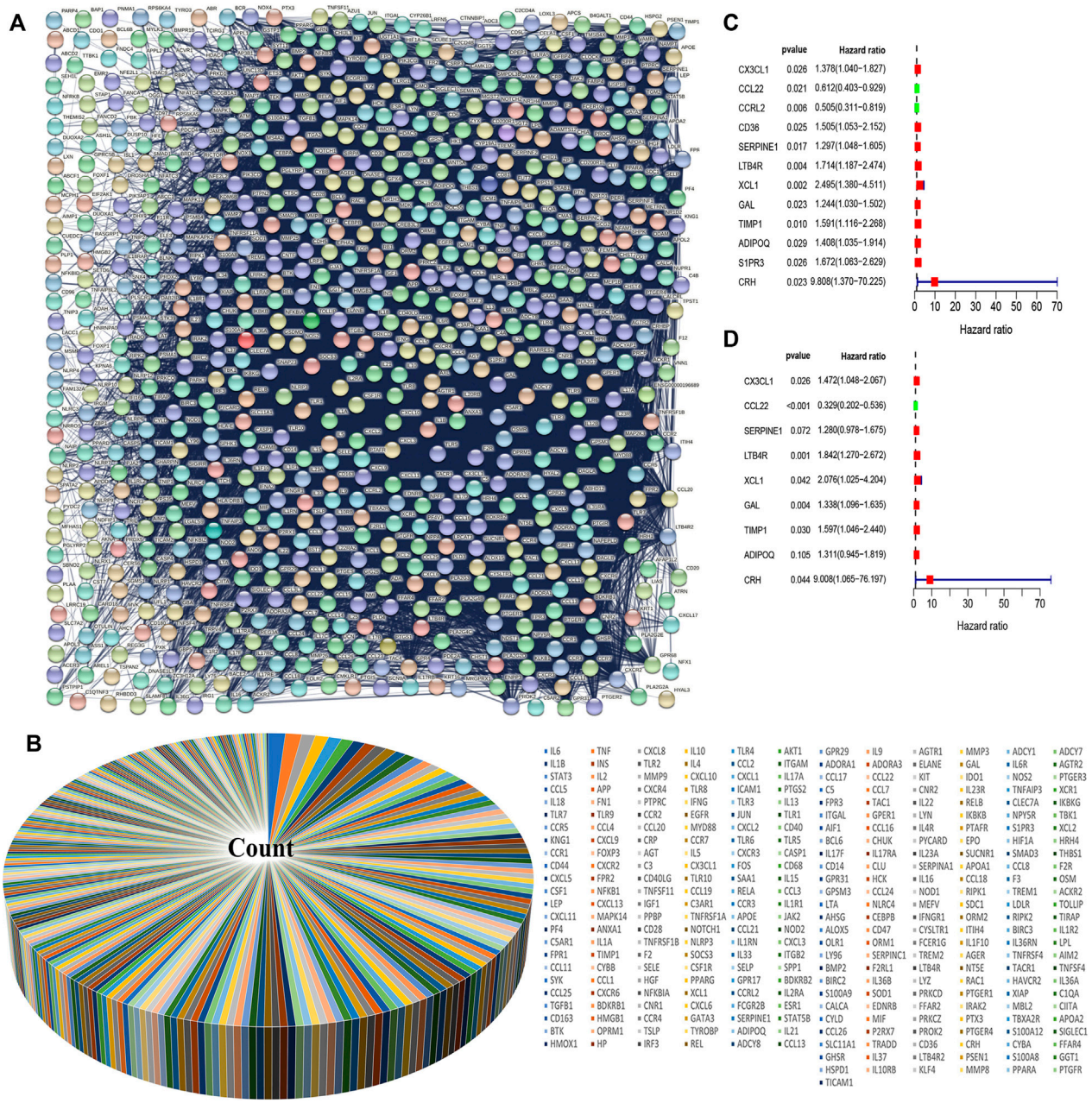


FIGURE 1 | Screening of inflammatory response genes and their influence on prognosis. **(A)** Protein interaction network of inflammatory response genes. **(B)** The 301 inflammatory response genes with the largest number of adjacent nodes have been represented by a pie chart. **(C)** The genes with *p* value <0.03 were screened by univariate Cox regression analysis. **(D)** Multivariate Cox regression analysis was performed to select the inflammatory response genes that can independently affect the prognosis.

we selected 301 genes which were the largest number of adjacent nodes for subsequent gene screening (**Figure 1B**). Next, we downloaded gene expression data and clinical information of intestinal tumors from the Cancer Genome Atlas (TCGA) database and extracted the expression levels of inflammatory response-related genes. Cox univariate analysis was used to screen out genes related to CRC prognosis (**Figure 1C**). Since there were several genes related to prognosis, we selected the genes with $p < 0.03$ for subsequent modeling. *CCL22* and *CCRL2* were identified as low risk genes, while *CX3CL1*, *CD36*, *SERPINE1*, *LTB4*, *XCL1*, *GAL*, *TIMP1*, *ADIPOQ*, *S1PR3*, and *CRH* were identified high-risk genes. We included these genes in subsequent Cox multivariate regression analysis. The final prognosis model was constructed using *CX3CL1*, *CCL22*, *SERPINE1*, *LTB4R*, *XCL1*, *GAL*, *TIMP1*, *ADIPOQ*, *CRH*, and *CCL22* which was a low-risk gene (**Figure 1D**).

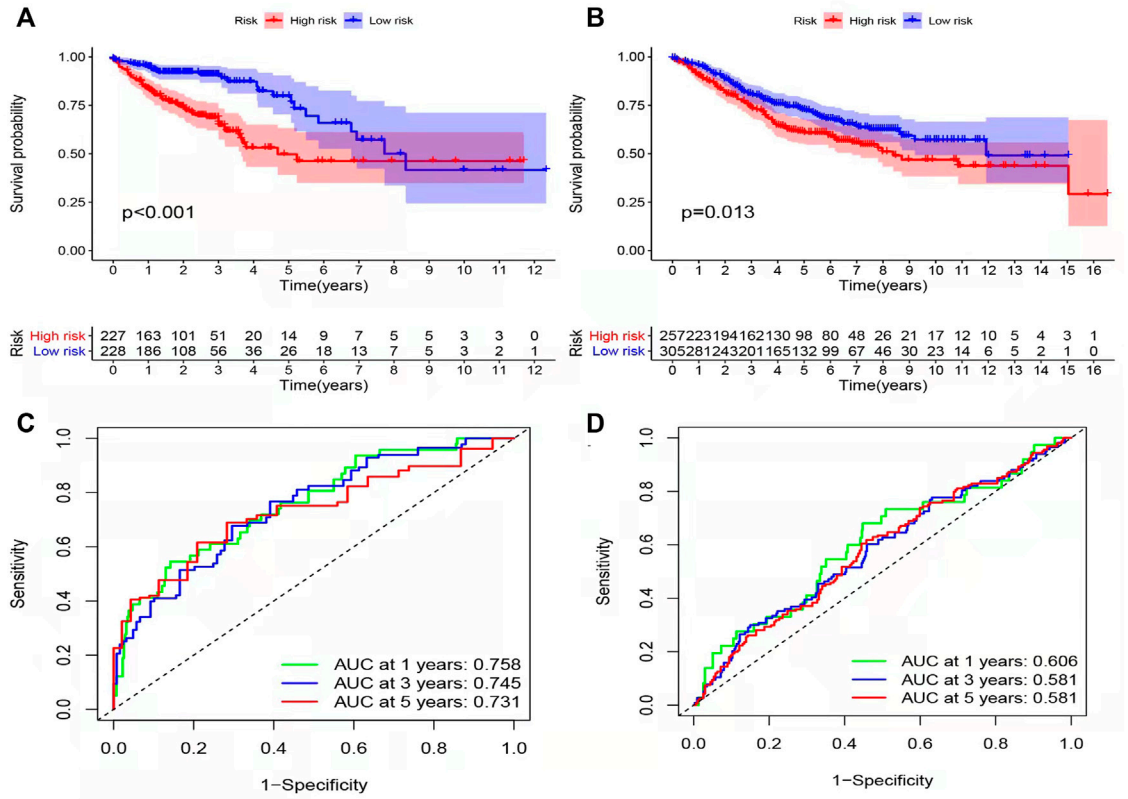


FIGURE 2 | Effect of the model on prognosis. **(A,B)** Patients with colorectal cancer were divided into high-risk group and low risk group using Kaplan Meier method (TCGA, GEO). The log-rank test was used to compare the survival time between the high-risk and low-risk groups (*p* values were less than 0.001 and equal to 0.013, respectively). **(C,D)** By using ROC curve to evaluate the accuracy of the prediction model. The area under the ROC curve of TCGA and GEO databases was greater than 0.5.

Effects of Inflammatory Response-Related Genes on Prognosis

Multivariate Cox regression analysis revealed nine inflammatory response genes which were related to the prognosis of colorectal cancer and were subsequently used to construct the prognosis model. The risk score for each patient was obtained by multiplying the amount of gene expression by the corresponding regression coefficients (0.3864, −1.111, 0.2468, 0.6110, 0.7304, 0.2915, 0.4683, 0.2706, and 2.198 for *CX3CL1*, *CCL22*, *SERPINE1*, *LTB4R*, *XCL1*, *GAL*, *TIMP1*, *ADIPOQ*, and *CRH*, respectively). The median risk score of the patients in the TCGA and GEO databases as the standard, the patients were divided into high and low-risk groups.

Results of survival analysis for the two groups of patients is shown in **Figures 2A, B**. Significant differences in the survival were observed between the high-risk and low-risk groups ($p < 0.05$). The area under the receiver operating characteristic (ROC) curve of the TCGA database and GEO database was greater than 0.05 (**Figures 2C, D**). However, the area under the curve did not increase with the increase in follow-up time for either the TCGA or GEO databases. This shows that the prediction model has some issues regarding the accuracy of prognosis and needs improvement. In order to observe the survival of the patients in the high and low risk groups more intuitively, we used a risk histogram to display the survival status of patients in the TCGA and GEO databases (**Figures 3A, B**). Patients in the high-risk group had lower survival rates than those in the low-risk group, indicating that our model can distinguish the high-risk group from the low-risk group. Next, we analyzed the relationship between the prognosis-related inflammatory response genes in the model (**Figures 3C, D**). The relationship between the risk score and survival rate of patients is illustrated by risk curves in **Figure 3** (E and F plotted the risk scores of the patients in the high and low risk groups, respectively). Patients in the medium-high risk group lived shorter lives than those in the low-risk group (**Figures 3G, H**), and with increasing time, the number of deaths also decreased. Finally, by using thermography, we evaluated the expression of the genes in the high and low risk groups in the model (**Figures 3I, J**).

Effect of Different Clinical Characteristics on the Prognosis of CRC

Different clinical characteristics have different effects on the prognosis of patients with CRC. Along with the risk score obtained from the model, we analyzed the influence of different clinical characteristics on the prognosis of patients

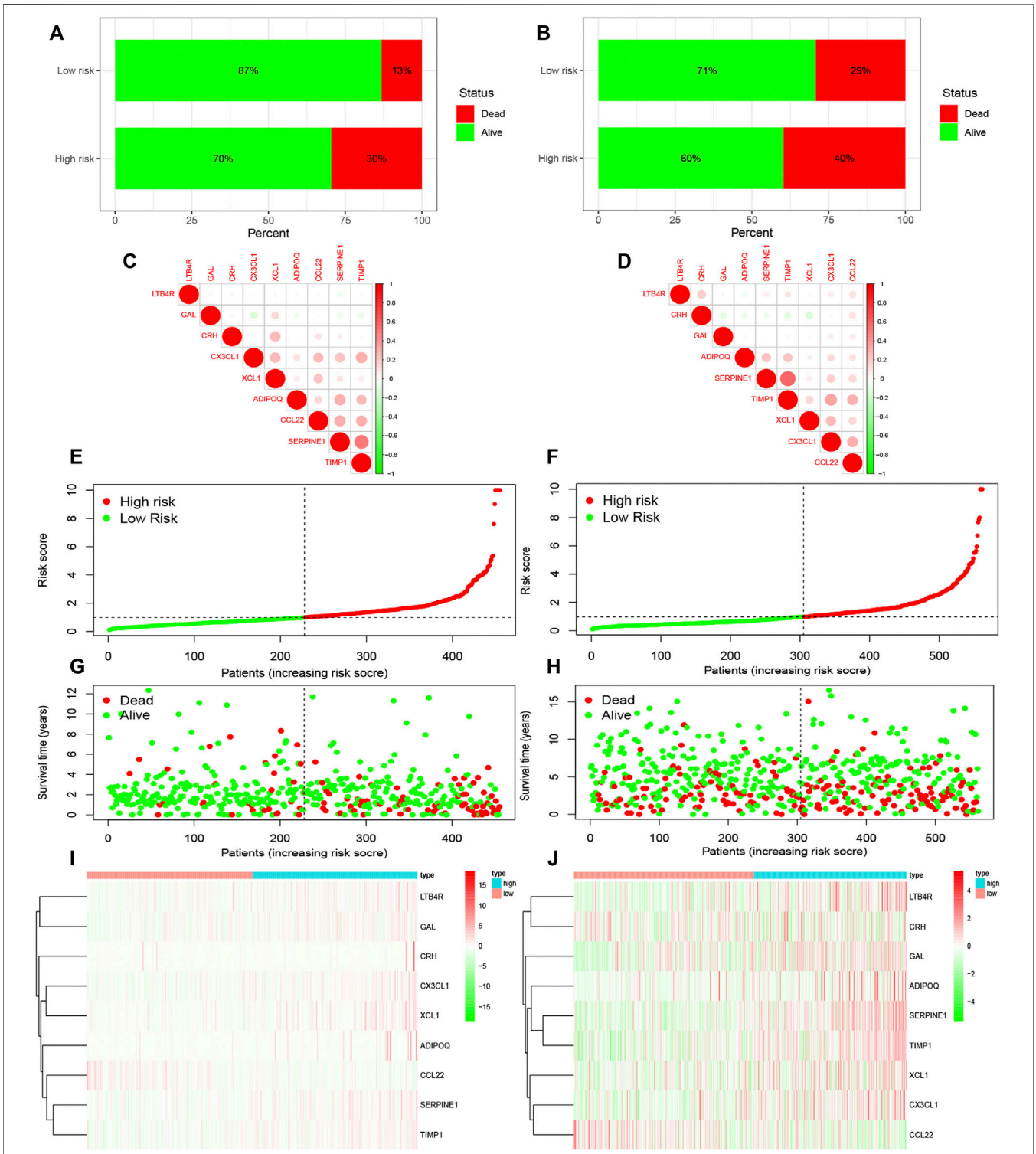


FIGURE 3 | Survival rate and prognosis of patients using the model, along with the expression of genes in the model in the high-risk and low-risk groups. **(A,B)** Survival rates in the TCGA and GEO databases. **(C,D)** Relationship between the genes in the model. **(E,F)** Risk scores of patients in the TCGA and GEO databases. **(G,H)** Survival time of patients in the high and low risk groups in the TCGA and GEO databases. **(I,J)** Expression of the genes in the model in the high and low risk groups of the TCGA and GEO databases.

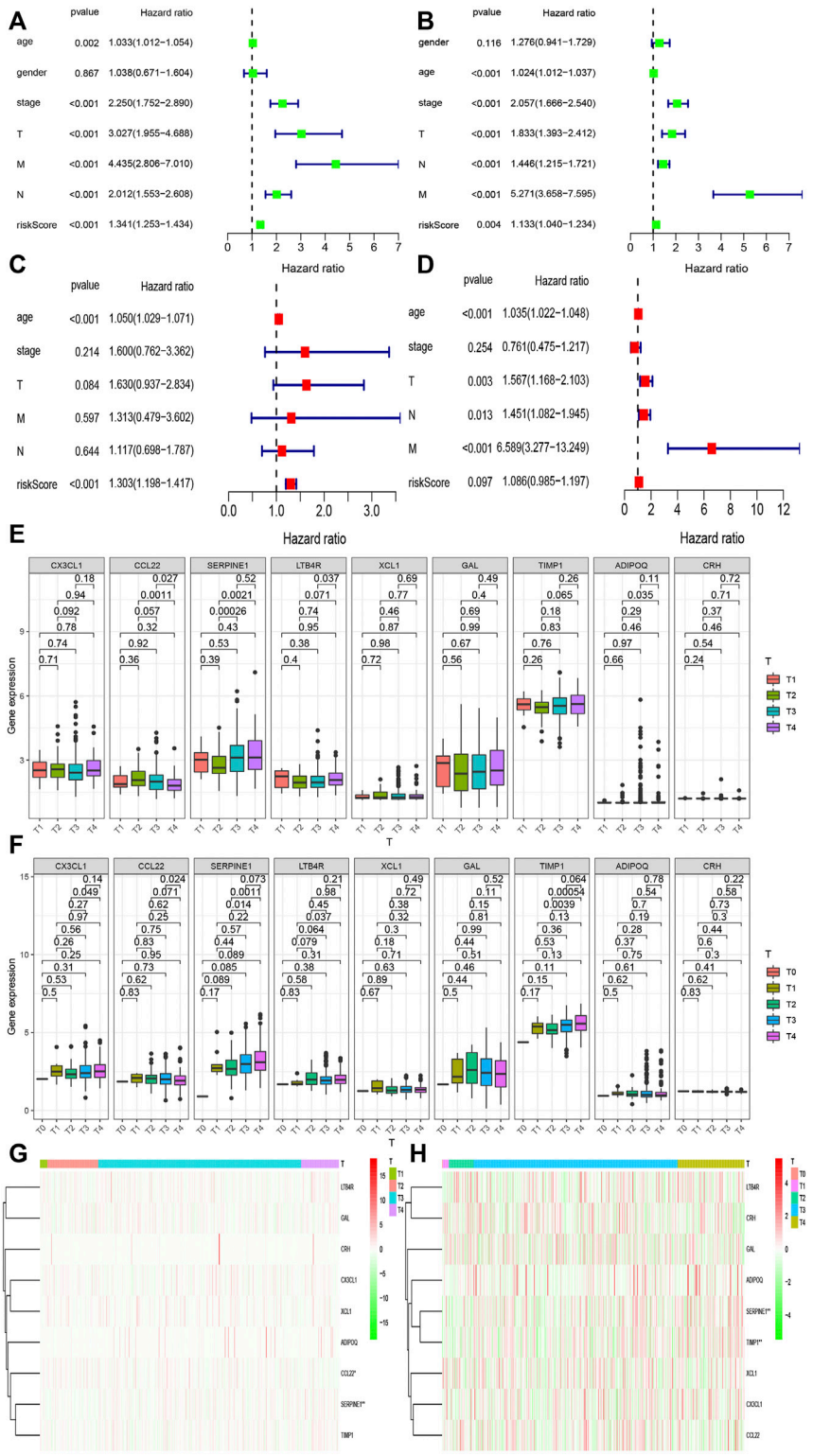


FIGURE 4 | Relationship between the risk model and different clinical traits. **(A,B)** Univariate prognostic analysis was performed on age, sex, T, N, M stage, and risk score of colorectal cancer patients in the TCGA and GEO databases. **(C,D)** Multivariate prognostic analysis was performed on age, sex, T, N, M stage, and risk score of colorectal cancer patients in the TCGA and GEO databases. **(E,F)** Expression of genes between among T stages in the TCGA and GEO databases. **(G,H)** Heat maps of the expression of inflammatory response genes among different T stages in models in the TCGA and GEO databases.

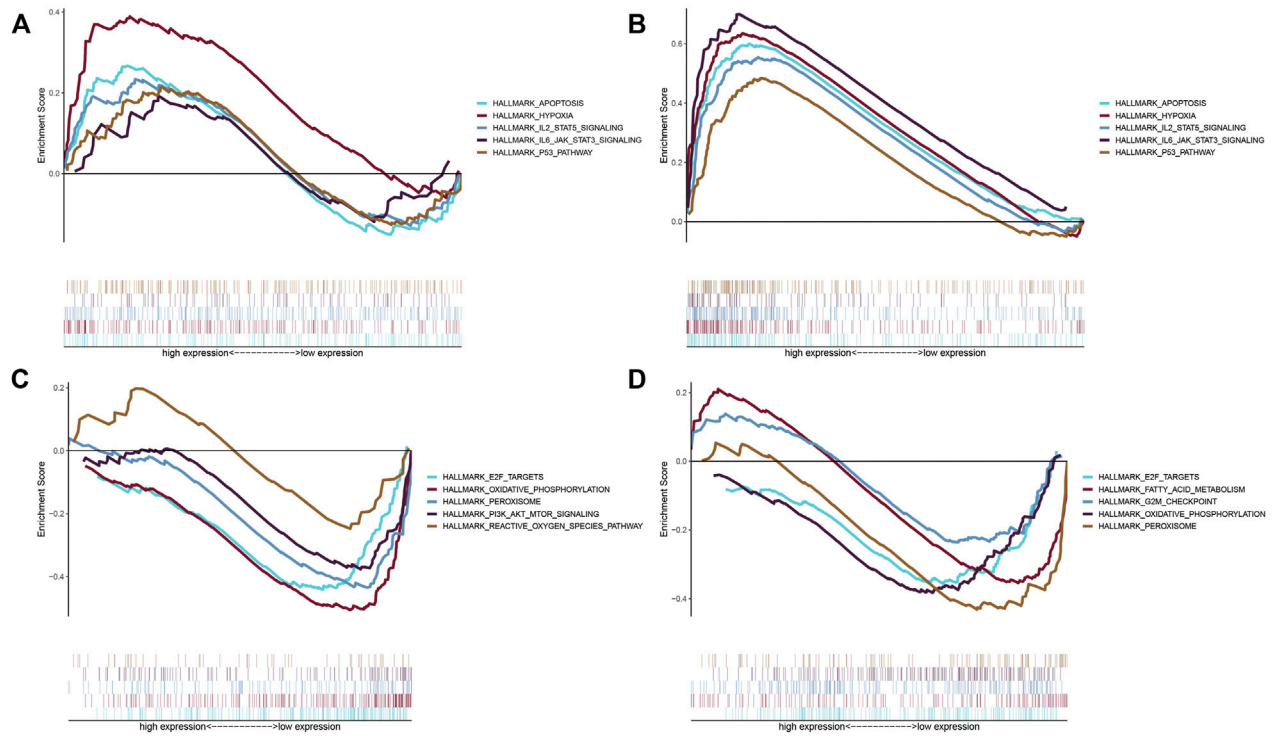


FIGURE 5 | Enrichment of pathways in the high-risk and low-risk groups. **(A)** Enriched gene sets in the hallmark gene collection by the high-risk score in the TCGA database. The lines with different colors represent different gene sets. The up-regulated genes are located near the origin of the coordinate on the left, while the down-regulated genes are located on the right side of the *x*-axis. Gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were statistically signi fi cant. Only a few major gene sets are shown. **(B)** Enriched gene sets in the hallmark collection by the high-risk score in the GEO database. Gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were statistically significant. Only a few major gene sets are shown. **(C)** Enriched gene sets in the hallmark collection by the low-risk score in the TCGA database. Only gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were considered statistically significant. Only a few major gene sets are shown. **(D)** Enriched gene sets in the hallmark collection by the low-risk score in the GEO database. Gene sets with NOM $p < 0.05$ and FDR $q < 0.06$ were statistically significant. Only a few major gene sets are shown.

with CRC. First, Cox univariate regression analysis was used to determine the association between clinical features and the prognosis of patients from the two databases (**Figures 4A, B**). Sex had no effect on the prognosis of patients, whereas other factors had an impact on the prognosis and were all high-risk factors. The risk score of the prediction model in the TCGA and GEO databases was also less than 0.05, which indicates that the risk score is related to prognosis of patients. Multivariate regression analysis of these factors showed that age, T, M, N, and risk score were all significant independent prognostic factors ($p < 0.05$) (**Figures 4C, D**). Next, we observed that there was a difference in the expression of inflammatory response-related genes between different T stages (**Figures 4E, F**). We found a significant difference in the expression of *SERPINE1* between different T stages ($p < 0.05$). **Figures 4G, H** are thermograms demonstrating the expression of the inflammatory response-related genes between different T stages in the two databases.

Enrichment of Inflammation Related Gene Pathways in the High and Low Risk Groups

There were also differences in the enrichment of inflammatory response-related genes between the high and low risk groups. In order to understand the enrichment of pathways, GSEA software was used to analyze the pathways in the two risk groups. The high-risk groups in the TCGA and GEO databases demonstrated enrichment of a large number of apoptotic, hypoxia, and inflammation related pathways, including apoptosis, hypoxia, IL-2-STAT5 signaling, IL-6-JAK-STAT3- signaling, and the P53-pathway (**Figures 5A,B**). The enriched pathways in the low-risk group were mostly related to oxidative phosphorylation and peroxidation including E2F-target, oxidative-phosphorylation, peroxisomes, PI3K-AKT-mTOR signaling, and reactive-oxygen-species pathways (**Figures 5C,D**).

Infiltration of Immune Cells

Inflammation often causes the infiltration of immune cells. Therefore, we studied the infiltration of immune cells in each risk group in the TCGA and GEO databases. **Figures 6A, B** respectively show the infiltration of immune cells in the high-risk groups and low-risk groups. In the TCGA database, the infiltration of 11 kinds of immune cells was significantly different between the high-risk and low-risk groups (**Figure 6C**, $p < 0.05$). In the GEO database, the infiltration of one kind of immune cell was different between the high-risk and low-risk groups (**Figure 6D**, $p < 0.05$). Activated dendritic cells

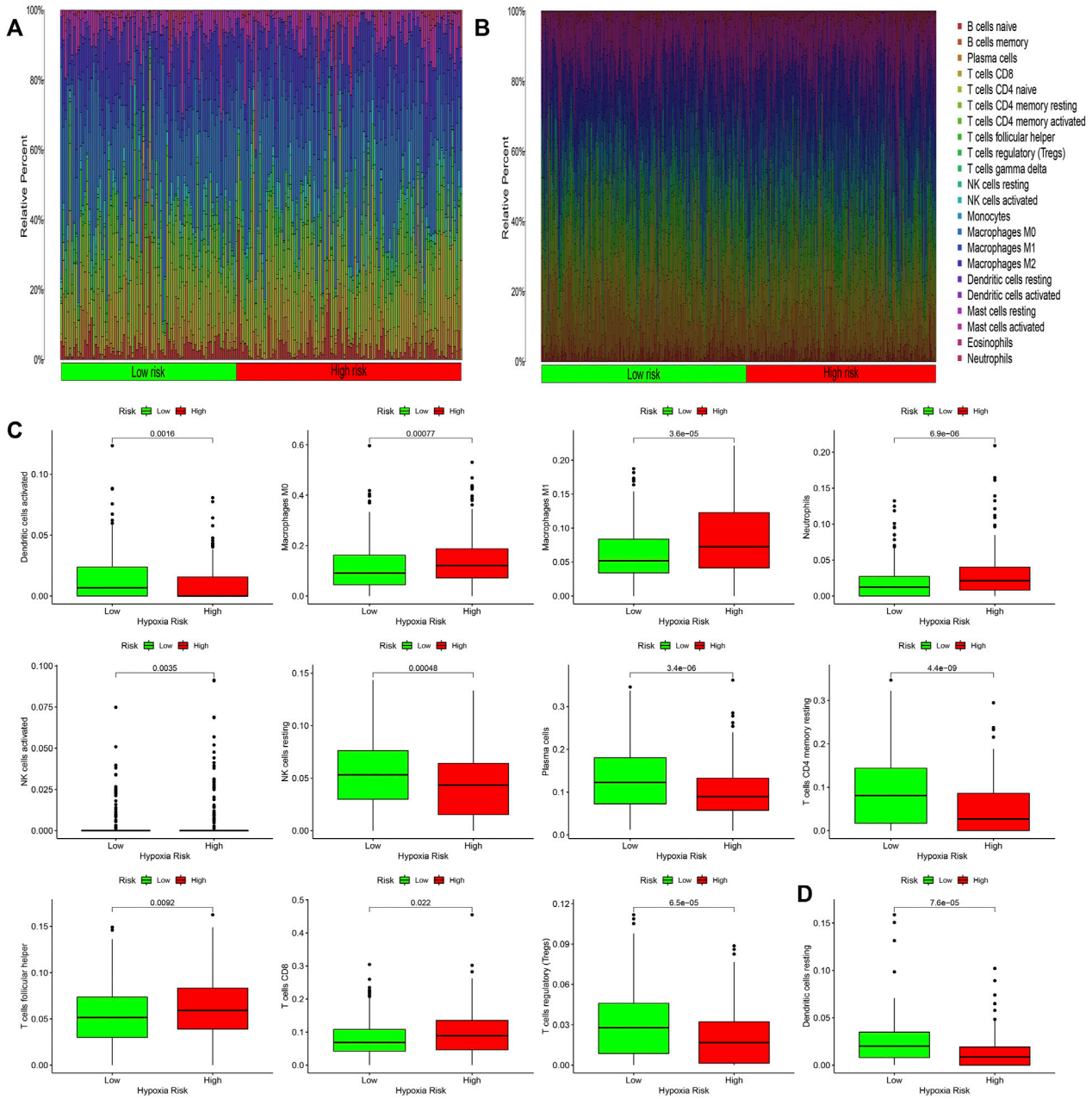


FIGURE 6 | Immune cell infiltration in high and low risk groups. **(A,B)** Thermography of inflammatory response-related gene risk scores and immune cell infiltration in the TCGA and GEO databases. **(C,D)** Infiltrating immune cells in the TCGA and GEO databases were significantly associated with the risk scores.

had different infiltration in the two databases. We downloaded immune related genes from the website of Tracking Tumor Immunophenotype and screened out the genes that regulate the activation of dendritic cells. We created heat maps (**Figures 7A, B**) to show the expression of genes regulating the infiltration of dendritic cells in the TCGA and GEO databases in the high and low risk groups. The expression of the *CCL21* gene was significantly different between the two databases ($p < 0.05$). A correlation curve between *CCL21* gene expression and the risk score was drawn, which revealed that *CCL21* gene expression was positively correlated with the risk score. Additionally, the expression of *CCL21* gene was different between the high-risk and low-risk groups (**Figures 7C, D**).

DISCUSSION

During tumor development, changes observed in the tumor sites resemble chronic inflammation, a process described as "a tumor is an unhealed wound" that promotes tumor survival

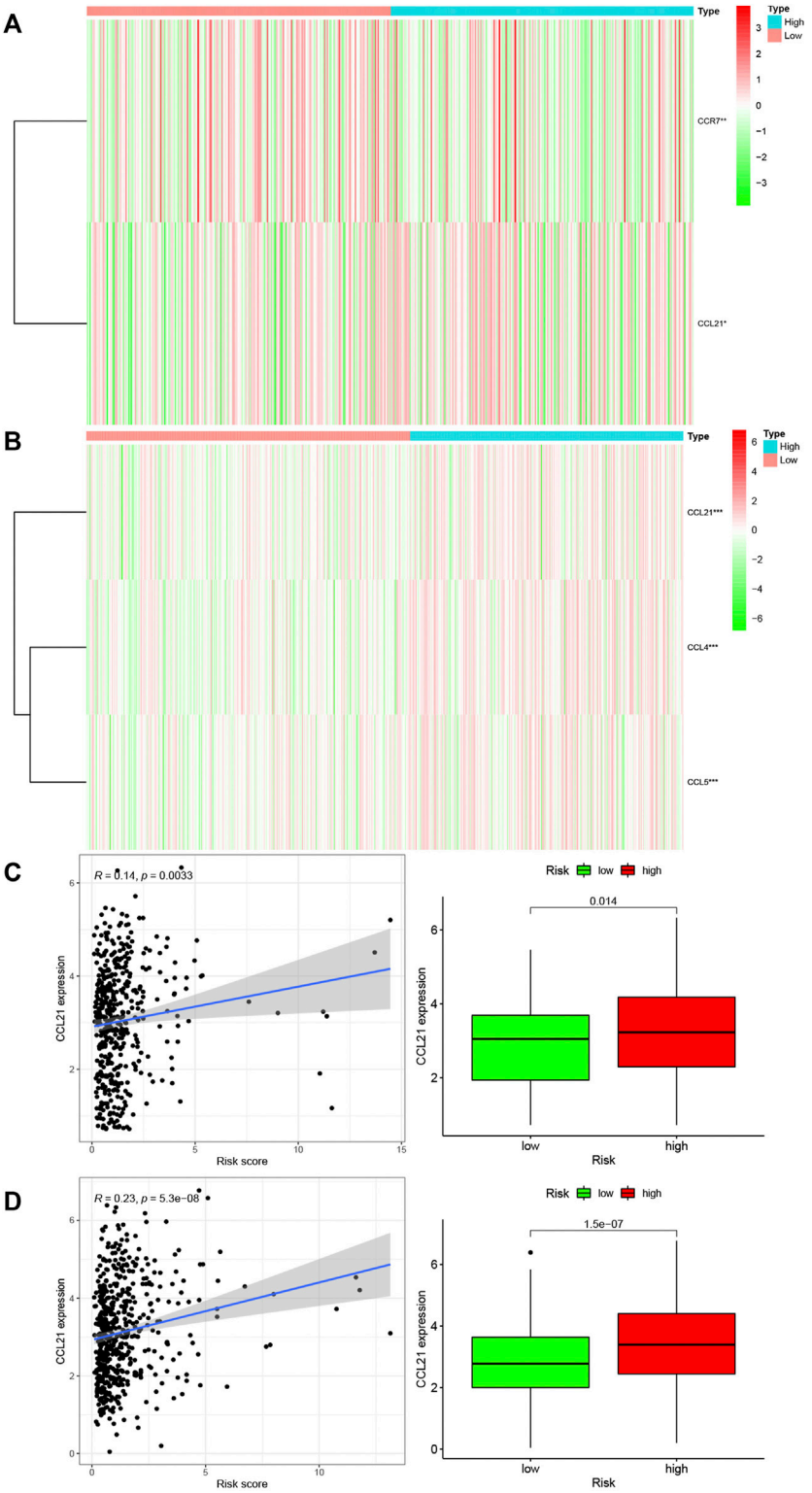


FIGURE 7 | Relationship between immune cell regulatory genes and risk scores of genes associated with inflammatory response. **(A,B)** Heat maps of the expression of genes significantly related to immune cell regulation in different risk groups in TCGA and GEO databases (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(C,D)** Correlation between CCL21 gene significantly associated with immune cell regulation and risk score in the TCGA and GEO databases. The blue line in the figure fitted the linear model between gene expression and risk score of genes related to inflammatory response, and Pearson's coefficient was used to test the correlation. Box chart showed the difference in CCL21 gene expression between groups with high and low risk of inflammatory response related genes ($p < 0.05$).

(Grivennikov et al., 2010). Studies have shown that chronic inflammation is the leading cause of many cancers in humans (Bernstein et al., 2001). Although inflammation can be used as a strategy against microbes, it is also thought to be a marker of cancer and plays a key role in tumorigenesis. Inflammatory response plays a decisive role in different stages of tumor progression, including initiation, promotion, malignant transformation, invasion, and metastasis (Grivennikov et al., 2010), and can promote carcinogenesis by inducing gene mutations, stimulating angiogenesis and cell proliferation, or inhibiting cell apoptosis (Rutter et al., 2004). Inflammation-related genes are involved in these processes and act on corresponding pathways or regulate immune cells. Our study screened the inflammatory response genes associated with CRC and identified the core genes, among which *CX3CL1*, *CCL22*, *SERPINE1*, *LTB4R*, *XCL1*, *GAL*, *TIMP1*, *ADIPOQ*, and *CRH* were closely related to the prognosis of patients. *CX3CL1* can induce the ERK pathway and cell proliferation, and also plays a specific tumor promoter role in breast cancer expressing *ERBB2* (Tardáguila et al., 2013). *CCL22* is a chemokine that is highly expressed in tumors, and promotes tumor growth, in addition to playing a role in tumor-related immunosuppression (Wiedemann et al., 2016). *SERPINE1* and *TIMP1* promote the migration and invasion of tumor cells (Song et al., 2016a; Klimczak-Bitner et al., 2016). *SERPINE1* may also promote the invasion and metastasis of colorectal cancer (Simone et al., 2015). *LTB4R* is a potent chemoattractant involved in inflammatory and immune responses to the paeoniflora-like signaling pathway (Wilson et al., 2014), which is involved in all inflammatory diseases (Peters-Golden and Henderson, 2007). *XCL1* promotes antitumor activity (Chou et al., 2020), and *XCL1* expression is also significantly related to the number of tumor infiltrating CD8+T cells as well as the expression of *PD-L1* in tumor cells (Tamura and Yoshihara, 2020). *GAL* methylation status may be an important marker for predicting clinical prognosis in patients who are with head and neck squamous cell carcinoma (Misawa et al., 2017). *ADIPOQ* gene play a role in chronic inflammation and cancer (Divella et al., 2017). *CRH* expression is associated with the advanced stage of ovarian cancer (Minas et al., 2007). It can be seen that these inflammatory response-related genes are closely related to tumors. Accordingly, we chose these nine genes to build an inflammatory model.

Inflammatory responses are associated with poor prognosis in a variety of tumors (Minas et al., 2007; Sano et al., 2018; Zhang et al., 2019). In order to further examine the relationship between the model and prognosis of patients, we evaluated the prognosis of patients by taking the product of the expression level and expression coefficient of the nine genes in the model in the two databases as the risk score. Significant differences in prognosis were found in the high-risk and low-risk groups, and patients in the low-risk group lived significantly longer than those in the high-risk group. However, in the process of using ROC curve to evaluate the accuracy of survival analysis, we found that the AUC value did not change much with the extension of time in the TCGA and GEO databases. This indicates that the accuracy of the ROC curve to evaluate survival was not ideal, which may be related to the following factors: First, the survival rate of CRC is 73.8% (70.0% for rectal cancer, 75.9% for colon cancer) (Fatemi et al., 2015), five- year survival rate is 68.4% (Kong et al., 2019), and average survival time is 142.17 ± 21.60 months (Fatemi et al., 2015). However, we evaluated 1-, 3-, and 5-years survival. Therefore, the survival situation could not be accurately reflected. Second, the sample was relatively small. Third, TCGA and GEO databases were selected for this study, and the predictive and prognostic accuracy of this model needs to be verified using multiple databases. Lastly, adjuvant therapies such as surgery and chemoradiotherapy also have an impact on the prognosis of patients (Fatemi et al., 2015). This study analyzed the impact of age, sex, T, M, N stage, and risk score on the prognosis of patients, and found that the risk score had a corresponding impact on the prognosis of patients in both databases. However, in the multivariate analysis, the *p* value of the risk score in the GEO database was >0.05, indicating that the risk score could not be used as an independent prognostic factor. On further examination, we found that all colorectal cancer types in the GEO database were adenocarcinomas, and the samples were all from France. Because of the limitations of tumor types and sample sources, the results may not accurately reflect the effect of inflammatory response on prognosis of colorectal cancer patients. Some patients had received chemotherapy, and the prognosis of CRC patients is related to the depth of tumor invasion, presence of lymph node metastasis (Haq et al., 2009), presence of other diseases, presence of venous or lymphatic invasion, tumor grade (Zlobec and Lugli, 2008), and genetic factors. These factors were not taken into account in this study, which is a limitation.

Furthermore, by using GSEA enrichment analysis to study the pathway enrichment in the high-risk groups and low-risk groups in the two databases. The study showed that the enriched pathways were mostly associated with hypoxia, inflammatory factors, and apoptosis.

Hypoxia and inflammation are closely related (Biddlestone et al., 2015). The inflammatory environment itself tends to be hypoxic (Watts and Walmsley, 2019), possibly because the metabolically active cells such as neutrophils migrating from the peripheral blood to the inflammatory tissue consume a large amount of energy (Borregaard and Herlin, 1982; Pollard and Borisy, 2003), and increase the oxygen demand (Rao and Suvas, 2019). Moreover, inflammation often leads to activation of the cellular hypoxia response pathways (Liu et al., 2014). Infection activates keratinocytes, macrophages, dendritic cells, and other cells, leading to the production of inflammatory cytokines (Wong and Goeddel, 1986; Mooney et al., 1990). Hypoxia also increases circulating proinflammatory cytokine levels (Song et al., 2016b). Tumor development has been shown to be associated with the inactivation of apoptosis (Xu et al., 2009). The downregulation of the tumor suppressor gene p53 can lead to reduced cell apoptosis and promote tumor growth (Bauer and Helfand, 2006), which are associated with many cancers (Rodrigues et al., 1990; Gasco et al., 2002). Some scientists also believe that apoptosis drives the proliferation and metastasis of tumor cells (Wang et al., 2013). Thus, there

is a close relationship between inflammatory response, inflammatory factors, and apoptosis.

The immune system plays a decisive role in the initial inflammatory response to infection and injury and is the main driver of the inflammatory protective response (Carrillo-Salinas et al., 2019). Therefore, the inflammatory response often leads to the infiltration of immune cells. In our study, the infiltration of 11 types of immune cells including activated dendritic cells, macrophages M0, macrophages M1, neutrophils, activated natural killer (NK) cells, NK cells, plasma cells, CD4 memory T cells, helper T cells, CD8 T cells, and regulatory T cells was significantly differently between the high-risk and low-risk groups in the TCGA database. However, only the infiltrates of activated dendritic cells in the GEO database were significantly different in the high and low risk groups. Subsequently, the genes regulating activated dendritic cells were screened, and it was found that the expression of *CCL2* gene was different in the groups with high and low risk in the TCGA and GEO databases. Other studies have also shown that the *CCL2* gene is associated with inflammatory responses. *CCL2* is a chemokine that attracts and activates monocytes (Conti and Rollins, 2004). *CCL2* plays a crucial role in tumor cell growth, metastasis, and host immune response (Zhuang et al., 2018). Additionally, *CCL2* has been shown to have both tumor stimulating and antitumor effects. Recent studies have suggested that *CCL2* plays a major role in tumor progression and metastasis (Li et al., 2013). *CCL2* can enhance the migration and invasion ability of prostate cancer cells (Natsagdorj and Izumi, 2019), as well as induce the invasion of liver cancer cells (Zhuang et al., 2018). High levels of *CCL2* expression in various types of tumors are also associated with poor prognosis (Yang et al., 2016); for example, the increased level of *CCL2* expression is related to poor prognosis in breast cancer patients (Lebrecht et al., 2004; Fang et al., 2015). Although many studies have elaborated the relationship between inflammatory response and colorectal cancer from the perspective of genetics and pharmacology, few articles have explored the relationship between inflammatory response-related genes and CRC at the genetic level. This study analyzed the relationship between inflammatory response-related genes and colorectal cancer at the genetic level, which can facilitate further research on colorectal cancer.

However, this study only carried out bioinformatics correlation analysis and did not explore the specific mechanism of inflammation response genes affecting prognosis. This study only proved that the risk scoring model established by us was related to prognosis of patients with colorectal cancer patients. Therefore, prospective studies, such as some basic and clinical studies, are needed to explore the specific mechanisms by which the genes we have identified interact with colorectal cancer.

In conclusion, inflammatory response plays a significant role in the prognosis of CRC patients and in the tumor immune microenvironment. Understanding the relationship between inflammatory response and immune cells is conducive to the faster application of effective immunotherapy for CRC treatment in the clinic, leading to an improvement in the prognosis of colorectal cancer patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/ and GEO (GSE39582) database.

AUTHOR CONTRIBUTIONS

Wrote the manuscript: YCW and LPZ. Conceived and designed the study: YW and NZ. Collected the literature: GHS, MQL, WDZ, and YLZ. All authors read and approved the final manuscript.

FUNDING

This work was supported by Finance Department of Jilin (2018SCZWSZX-039) and Health Department of Jilin (2015Q011) and Finance Department of Jilin (JLSWSRCZX 2020-083) and The National Natural Science Foundation of China (81772055) and Science and Technology Department of Jilin Province (20170623092TC-11) and Key research project of Science and Technology Department of Jilin Province (20130206063YY) and Education Department of Jilin Province (JJKH20211209KJ) and Education Department of Jilin Province (JJKH20221064KJ) and Health Commission of Jilin Province (2021JC011).

ACKNOWLEDGMENTS

We would like to acknowledge the reviewers for their helpful comments on this study.

REFERENCES

Bauer, J. H., and Helfand, S. L. (2006). New Tricks of an Old Molecule: Lifespan Regulation by P53. *Aging Cell* 5 (5), 437–440. doi:10.1111/j.1474-9726.2006.00228.x

Bernstein, C. N., Blanchard, J. F., Kliewer, E., and Wajda, A. (2001). Cancer Risk in Patients with Inflammatory Bowel Disease. *Cancer* 91 (4), 854–862. doi:10.1002/1097-0142(20010215)91:4<854::aid-cncr1073>3.0.co;2-z

Biddlestone, J., Bandarra, D., and Rocha, S. (2015). The Role of Hypoxia in Inflammatory Disease (Review). *Int. J. Mol. Med.* 35 (4), 859–869. doi:10.3892/ijmm.2015.2079

Borregaard, N., and Herlin, T. (1982). Energy Metabolism of Human Neutrophils during Phagocytosis. *J. Clin. Invest.* 70 (3), 550–557. doi:10.1172/jci110647

Brennan, C. A., and Garrett, W. S. (2016). Gut Microbiota, Inflammation, and Colorectal Cancer. *Annu. Rev. Microbiol.* 70, 395–411. doi:10.1146/annurev-micro-102215-095513

Carrillo-Salinas, F. J., Ngwenyama, N., Anastasiou, M., Kaur, K., and Alcaide, P. (2019). Heart Inflammation. *Am. J. Pathol.* 189 (8), 1482–1494. doi:10.1016/j.ajpath.2019.04.009

Chou, W. C., Hsiung, C. N., Chen, W. T., Tseng, L. M., Wang, H. C., Chu, H. W., et al. (2020). A Functional Variant Near XCL1 Gene Improves Breast Cancer Survival via Promoting Cancer Immunity. *Int. J. Cancer* 146 (8), 2182–2193. doi:10.1002/ijc.32855

Conti, I., and Rollins, B. J. (2004). CCL2 (Monocyte Chemoattractant Protein-1) and Cancer. *Semin. Cancer Biol.* 14 (3), 149–154. doi:10.1016/j.semcancer.2003.10.009

Divella, R., Daniele, A., Mazzocca, A., Abbate, I., Casamassima, P., Caliandro, C., et al. (2017). ADIPOQ Rs266729 G/C Gene Polymorphism and Plasmatic Adipocytokines Connect Metabolic Syndrome to Colorectal Cancer. *J. Cancer* 8 (6), 1000–1008. doi:10.7150/jca.17515

Drewes, J. L., Housseau, F., and Sears, C. L. (2016). Sporadic Colorectal Cancer: Microbial Contributors to Disease Prevention, Development and Therapy. *Br. J. Cancer* 115 (3), 273–280. doi:10.1038/bjc.2016.189

Esteva, F. J., Hubbard-Lucey, V. M., Tang, J., and Pusztai, L. (2019). Immunotherapy and Targeted Therapy Combinations in Metastatic Breast Cancer. *Lancet Oncol.* 20 (3), e175–e186. doi:10.1016/s1470-2045(19)30026-9

Fang, W. B., Yao, M., Jokar, I., Alhakamy, N., Berkland, C., Chen, J., et al. (2015). The CCL2 Chemokine Is a Negative Regulator of Autophagy and Necrosis in Luminal B Breast Cancer Cells. *Breast Cancer Res. Treat.* 150 (2), 309–320. doi:10.1007/s10549-015-3324-4

Fatemi, S. R., Pourhoseingholi, M. A., Asadi, F., Vahedi, M., Pasha, S., Alizadeh, L., et al. (2015). Recurrence and Five Year Survival in Colorectal Cancer Patients after Surgery. *Iran J. Cancer Preven* 8 (4), e3439. doi:10.17795/ijcp.3439

Ferrari, S. M., Fallahi, P., Galdiero, M. R., Ruffilli, I., Elia, G., and Ragusa, F., (2019). Immune and Inflammatory Cells in Thyroid Cancer Microenvironment. *Ijms* 20 (18), 4413. doi:10.3390/ijms20184413

Gasco, M., Shami, S., and Crook, T. (2002). The P53 Pathway in Breast Cancer. *Breast Cancer Res.* 4 (2), 70–76. doi:10.1186/bcr426

Grivennikov, S. I., Greten, F. R., and Karin, M. (2010). Immunity, Inflammation, and Cancer. *Cell* 140 (6), 883–899. doi:10.1016/j.cell.2010.01.025

Haq, A. I., Schneeweiss, J., Kalsi, V., and Arya, M. (2009). The Dukes Staging System: a Cornerstone in the Clinical Management of Colorectal Cancer. *Lancet Oncol.* 10 (11), 1128. doi:10.1016/s1470-2045(09)70157-3

Hooper, L. V., Littman, D. R., and Macpherson, A. J. (2012). Interactions between the Microbiota and the Immune System. *Science* 336 (6086), 1268–1273. doi:10.1126/science.1223490

Klimczak-Bitner, A. A., Kordek, R., Bitner, J., Musiał, J., and Szemraj, J. (2016). Expression of MMP9, SERPINE1 and miR-134 as Prognostic Factors in Esophageal Cancer. *Oncol. Lett.* 12 (5), 4133–4138. doi:10.3892/ol.2016.5211

Kong, L., Peng, J., Li, J., Wang, F., Li, C., Ding, P., et al. (2019). Prolonged Surveillance of Colorectal Cancer Patients after Curative Surgeries beyond Five Years of Follow-Up. *Ann. Transl. Med.* 7 (21), 608. doi:10.21037/atm.2019.10.39

Lebrecht, A., Grimm, C., Lantzsch, T., Ludwig, E., Hefler, L., Ulbrich, E., et al. (2004). Monocyte Chemoattractant Protein-1 Serum Levels in Patients with Breast Cancer. *Tumor Biol.* 25 (1-2), 14–17. doi:10.1159/000077718

Li, M., Knight, D. A., Snyder, M. J., and Stewart, T. J. (2013). A Role for CCL2 in Both Tumor Progression and Immunosurveillance. *Oncoimmunology* 2 (7), e25474. doi:10.4161/onci.25474

Liu, H., Wang, Z., Yu, S., and Xu, J. (2014). Proteasomal Degradation of O-GlcNAc Transferase Elevates Hypoxia-Induced Vascular Endothelial Inflammatory Response†. *Cardiovasc. Res.* 103 (1), 131–139. doi:10.1093/cvr/cvu116

Minas, V., Rolaki, A., Kalantaridou, S. N., Sidiropoulos, J., Mitrou, S., Petsas, G., et al. (2007). Intratumoral CRH Modulates Immuno-Escape of Ovarian Cancer Cells through FasL Regulation. *Br. J. Cancer* 97 (5), 637–645. doi:10.1038/sj.bjc.6603918

Misawa, K., Mochizuki, D., Endo, S., Mima, M., Misawa, Y., Imai, A., et al. (2017). Site-specific Methylation Patterns of theGALandGALR1/2genes in Head and Neck Cancer: Potential Utility as Biomarkers for Prognosis. *Mol. Carcinog.* 56 (3), 1107–1116. doi:10.1002/mc.22577

Mooney, D. P., O'REILLY, M., and Gamelli, R. L. (1990). Tumor Necrosis Factor and Wound Healing. *Ann. Surg.* 211 (2), 124–129. doi:10.1097/00000658-199002000-00002

Natsagdorj, A., Izumi, K., Hiratsuka, K., Machioka, K., Iwamoto, H., Naito, R., et al. (2019). CCL2 Induces Resistance to the Antiproliferative Effect of Cabazitaxel in Prostate Cancer Cells. *Cancer Sci.* 110 (1), 279–288. doi:10.1111/cas.13876

Peters-Golden, M., and Henderson, W. R., Jr. (2007). Leukotrienes. *N. Engl. J. Med.* 357 (18), 1841–1854. doi:10.1056/NEJMra071371

Pollard, T. D., and Borisy, G. G. (2003). Cellular Motility Driven by Assembly and Disassembly of Actin Filaments. *Cell* 112 (4), 453–465. doi:10.1016/s0092-8674(03)00120-x

Rao, P., and Suvas, S. (2019). Development of Inflammatory Hypoxia and Prevalence of Glycolytic Metabolism in Progressing Herpes Stromal Keratitis Lesions. *J.I.* 202 (2), 514–526. doi:10.4049/jimmunol.1800422

Rizzo, A., Pallone, F., Monteleone, G., and Fantini, M. C. (2011). Intestinal Inflammation and Colorectal Cancer: a Double-Edged Sword? *World J. Gastroenterol.* 17 (26), 3092–3100. doi:10.3748/wjg.v17.i26.3092

Rodrigues, N. R., Rowan, A., Smith, M. E., Kerr, I. B., Bodmer, W. F., Gannon, J. V., et al. (1990). p53 Mutations in Colorectal Cancer. *Proc. Natl. Acad. Sci. U.S.A.* 87 (19), 7555–7559. doi:10.1073/pnas.87.19.7555

Rutter, M., Saunders, B., Wilkinson, K., Rumbles, S., Schofield, G., Kamm, M., et al. (2004). Severity of Inflammation Is a Risk Factor for Colorectal Neoplasia in Ulcerative Colitis. *Gastroenterology* 126 (2), 451–459. doi:10.1053/j.gastro.2003.11.010

Sano, Y., Kogashiwa, Y., Araki, R., Enoki, Y., Ikeda, T., Yoda, T., et al. (2018). Correlation of Inflammatory Markers, Survival, and COX2 Expression in Oral Cancer and Implications for Prognosis. *Otolaryngol. Head Neck Surg.* 158 (4), 667–676. doi:10.1177/0194599817745284

Simone, T. M., Longmate, W. M., Law, B. K., and Higgins, P. J. (2015). Targeted Inhibition of PAI-1 Activity Impairs Epithelial Migration and Wound Closure Following Cutaneous Injury. *Adv. Wound Care* 4 (6), 321–328. doi:10.1089/wound.2014.0611

Song, G., Xu, S., Zhang, H., Wang, Y., Xiao, C., Jiang, T., et al. (2016a). TIMP1 Is a Prognostic Marker for the Progression and Metastasis of colon Cancer through FAK-Pi3k/AKT and MAPK Pathway. *J. Exp. Clin. Cancer Res.* 35 (1), 148. doi:10.1186/s13046-016-0427-7

Song, T.-T., Bi, Y.-H., Gao, Y.-Q., Huang, R., Hao, K., Xu, G., et al. (2016b). Systemic Pro-inflammatory Response Facilitates the Development of Cerebral Edema during Short Hypoxia. *J. Neuroinflammation* 13 (1), 63. doi:10.1186/s12974-016-0528-4

Stoffel, E. M., and Yurgelun, M. B. (2016). Genetic Predisposition to Colorectal Cancer: Implications for Treatment and Prevention. *Semin. Oncol.* 43 (5), 536–542. doi:10.1053/j.seminoncol.2016.08.002

Strasser, K., Birnleitner, H., Beer, A., Pils, D., Gerner, M. C., Schmetterer, K. G., et al. (2019). Immunological Differences between Colorectal Cancer and normal Mucosa Uncover a Prognostically Relevant Immune Cell Profile. *OncoImmunology* 8 (2), e1537693. doi:10.1080/2162402x.2018.1537693

Tamura, R., Yoshihara, K., Nakaoka, H., Yachida, N., Yamaguchi, M., Suda, K., et al. (2020). XCL1 Expression Correlates with CD8-Positive T Cells Infiltration and PD-L1 Expression in Squamous Cell Carcinoma Arising from Mature Cystic Teratoma of the Ovary. *Oncogene* 39 (17), 3541–3554. doi:10.1038/s41388-020-1237-0

Tardáguila, M., Mira, E., García-Cabezas, M. A., Feijoo, A. M., Quintela-Fandino, M., Azcoitia, I., et al. (2013). CX3CL1 Promotes Breast Cancer via Transactivation of the EGF Pathway. *Cancer Res.* 73 (14), 4461–4473. doi:10.1158/0008-5472.can-12-3828

Wang, R.-A., Li, Q.-L., Li, Z.-S., Zheng, P.-J., Zhang, H.-Z., Huang, X.-F., et al. (2013). Apoptosis Drives Cancer Cells Proliferate and Metastasize. *J. Cel. Mol. Med.* 17 (1), 205–211. doi:10.1111/j.1582-4934.2012.01663.x

Watts, E. R., and Walmsley, S. R. (2019). Inflammation and Hypoxia: HIF and PHD Isoform Selectivity. *Trends Mol. Med.* 25 (1), 33–46. doi:10.1016/j.molmed.2018.10.006

Wiedemann, G. M., Knott, M. M. L., Vetter, V. K., Rapp, M., Haubner, S., Fesseler, J., et al. (2016). Cancer Cell-Derived IL-1α Induces CCL22 and the Recruitment of Regulatory T Cells. *Oncoimmunology* 5 (9), e1175794. doi:10.1080/2162402x.2016.1175794

Wilson, G. A., Butcher, L. M., Foster, H. R., Feber, A., Roos, C., Walter, L., et al. (2014). Human-specific Epigenetic Variation in the Immunological Leukotriene B4 Receptor (LTB4R/BLT1) Implicated in Common Inflammatory Diseases. *Genome Med.* 6 (3), 19. doi:10.1186/gm536

Wong, G. H. W., and Goeddel, D. V. (1986). Tumour Necrosis Factors α and β Inhibit Virus Replication and Synergize with Interferons. *Nature* 323 (6091), 819–822. doi:10.1038/323819a0

Xu, C.-X., Jin, H., and Cho, M.-H. (2009). Apoptosis and Apoptosis-Based Therapy in Lung Cancer. *Acamc* 9 (9), 952–957. doi:10.2174/187152009789377682

Yang, Y., Zhai, C., Chang, Y., Zhou, L., Shi, T., Tan, C., et al. (2016). High Expression of Chemokine CCL2 Is Associated with Recurrence after Surgery in clear-cell Renal Cell Carcinoma. *Urol. Oncol. Semin. Original Invest.* 34 (5), e19e219–238. doi:10.1016/j.urolonc.2015.11.026

Zhang, K., Hua, Y.-Q., Wang, D., Chen, L.-Y., Wu, C.-J., Chen, Z., et al. (2019). Systemic Immune-Inflammation index Predicts Prognosis of Patients with Advanced Pancreatic Cancer. *J. Transl Med.* 17 (1), 30. doi:10.1186/s12967-019-1782-x

Zhuang, H., Cao, G., Kou, C., and Liu, T. (2018). CCL2/CCR2 axis Induces Hepatocellular Carcinoma Invasion and Epithelial-Mesenchymal Transition Inï¿½vitro through Activation of the Hedgehog Pathway. *Oncol. Rep.* 39 (1), 21–30. doi:10.3892/or.2017.6069

Zlobec, I., and Lugli, A. (2008). Prognostic and Predictive Factors in Colorectal Cancer. *Postgrad. Med. J.* 84 (5), 403–411. doi:10.1136/jcp.2007.054858

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, Zhang, Shi, Liu, Zhao, Zhang, Wang 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.

frontiers | Frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fmolb.2022.874475

Artificial Intelligence Assisting the Early Detection of Active Pulmonary Tuberculosis From Chest X-Rays: A Population-Based Study

Mayidili Nijiati[1]†, Jie Ma[2†], Chuling Hu[3†], Abudouresuli Tuersun[1], Abudoukeyoumujiang Abulizi[1], Abudoureyimu Kelimu[4], Dongyu Zhang[2], Guanbin Li[2]* and Xiaoguang Zou[5]**

[1]Department of Radiology, The First People's Hospital of Kashi Prefecture, Kashi, China, [2]School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China, [3]Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [4]Department of Radiology, Kashi Area Tuberculosis Control Center, Kashi, China, [5]Clinical Medical Research Center, The First People's Hospital of Kashi Prefecture, Kashi, China

As a major infectious disease, *tuberculosis* (TB) still poses a threat to people's health in China. As a triage test for TB, reading chest radiography with traditional approach ends up with high inter-radiologist and intra-radiologist variability, moderate specificity and a waste of time and medical resources. Thus, this study established a deep convolutional neural network (DCNN) based artificial intelligence (AI) algorithm, aiming at diagnosing TB on posteroanterior chest X-ray photographs in an effective and accurate way. Altogether, 5,000 patients with TB and 4,628 patients without TB were included in the study, totaling to 9,628 chest X-ray photographs analyzed. Splitting the radiographs into a training set (80.4%) and a testing set (19.6%), three different DCNN algorithms, including ResNet, VGG, and AlexNet, were trained to classify the chest radiographs as images of pulmonary TB or without TB. Both the diagnostic accuracy and the area under the receiver operating characteristic curve were used to evaluate the performance of the three AI diagnosis models. Reaching an accuracy of 96.73% and marking the precise TB regions on the radiographs, ResNet algorithm-based AI outperformed the rest models and showed excellent diagnostic ability in different clinical subgroups in the stratification analysis. In summary, the ResNet algorithm-based AI diagnosis system provided accurate TB diagnosis, which could have broad prospects in clinical application for TB diagnosis, especially in poor regions with high TB incidence.

Keywords: tuberculosis, chest radiograph, machine learning, artificial intelligence, deep convolutional neural network

Abbreviations: AI, artificial intelligence; AUC, area under the curve; CAMs, class activation maps; DCNN, deep convolutional neural network; ROC, receiver operating characteristic; TB, *tuberculosis*.

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Xiaohan Xing,
City University of Hong Kong, Hong Kong SAR, China
Jun Cheng,
Shenzhen University, China

***Correspondence:**
Mayidili Nijiati
1376906729@qq.com
Guanbin Li
liguanbin@mail.sysu.edu.cn
Xiaoguang Zou
zxgks@163.com

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

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *12 February 2022*
Accepted: *08 March 2022*
Published: *08 April 2022*

Citation:
Nijiati M, Ma J, Hu C, Tuersun A, Abulizi A, Kelimu A, Zhang D, Li G and Zou X (2022) Artificial Intelligence Assisting the Early Detection of Active Pulmonary Tuberculosis From Chest X-Rays: A Population-Based Study. Front. Mol. Biosci. 9:874475. doi: 10.3389/fmolb.2022.874475

INTRODUCTION

Causing by *Mycobacterium tuberculosis* infection, pulmonary *tuberculosis* (TB) is a kind of dangerous airborne chronic respiratory infectious disease (Ragonnet et al., 2019). TB still remains a major problem of disease control and poses a threat to the health of the public in China (Tusun et al., 2021). The incidence of TB in Kashi was 250.4/100,000 in 2020, reaching 4.3 times the national average incidence. TB epidemic remains severe, especially in 12 counties/cities of Kashi prefecture, Xinjiang Uygur Autonomous Region (Tusun et al., 2021).

A triage test using chest radiography is utilized for patients with typical symptoms for TB or TB-related risk factors (Qin et al., 2021a). Both the shortage of experienced radiologist and the high inter-radiologist and intra-radiologist variability have been affecting the performance and generalizability of the chest radiography, especially in places with a high incidence of TB and without access to high quality medical service (Organization, 2016). However, during the last 10 years, the artificial intelligence (AI) aided diagnostics systems have been developing and evolving at an unprecedented pace, leading to its deployment and usage in clinical settings, and many medical image analyzing AI algorithms, which were based on deep learning and deep convolutional neural networks (DCNNs), were being utilized for radiographs reading at the same time (Rajpurkar et al., 2020; Qin et al., 2021b). Such deep learning and DCNN algorithms are able to distinguish the features and characteristics of the TB-related abnormalities in the chest X -ray photographs. Considering the great improvement of AI-assisted TB diagnosis, computer aided TB screening software is a better substitute for physicians in digital chest radiographs reading and analyzing, which was recommended by the updated guidelines of the World Health Organization (WHO) in March 2021 (Organization, 2021). However, there remains uncertain about what kind of algorithm or AI should be developed and put into clinical practice, since the WHO didn't give detailed recommendations for specific products (Qin et al., 2021a).

To date, the majority of AI algorithms for TB diagnosis have been based on small groups of individuals. Considering that a large sample for training would further improve the performance of the AI algorithm, the deep convolutional neural network (DCNN), a kind of deep learning approach, has been widely utilized for analyzing medical images. Thus, this study explored the TB diagnosis ability of three kinds of DCNNs (Resnet, VGG, and AlexNet algorithms) based on chest X-rays of 10,000 individuals.

MATERIALS AND METHODS

Study Setting and Population

In this retrospective study, we trained convolutional neural network-based AI algorithms to read chest X-rays for pulmonary TB diagnosis. The workflow of the study was showed in **Figure 1**. In total, 9628 X-ray images and corresponding clinical information were collected from individuals with and without TB in Kashgar, Xinjiang, China, between 2019 and 2020 (**Table 1**). The included cases were aged ≥15 years and underwent X-ray analysis. TB cases were diagnosed by experienced physicians based on the symptoms and the results of multiple tests and radiological examinations, including sputum culture or smear tests, Xpert tests, chest X-ray films, interferon gamma release assays and tuberculin skin tests and so on. In total, 5,000 images of TB cases and 4,628 images of non-TB cases were collected with privacy information removed and split into the training (n = 7,703) and testing sets (n = 1925) (**Table 1**).

X-Rays Images Preprocessing

As a tool widely used in medical image semantic segmentation, U-Net has been applied to extract semantic information and generate segmentation results (Ronneberger et al., 2015; Zunair and Ben Hamza, 2021). In order to focus on pulmonary TB-affected regions that appeared inside the lungs, U-Net was used for the lung segment before TB classification. After image cropping and resizing, the lung region images served as an input for the classification convolutional neural network (**Figure 2**).

Development of AI Algorithms

For the classification network, we used ResNet34 (He et al., 2016), VGG (Simonyan and Zisserman, 2014) and AlexNet

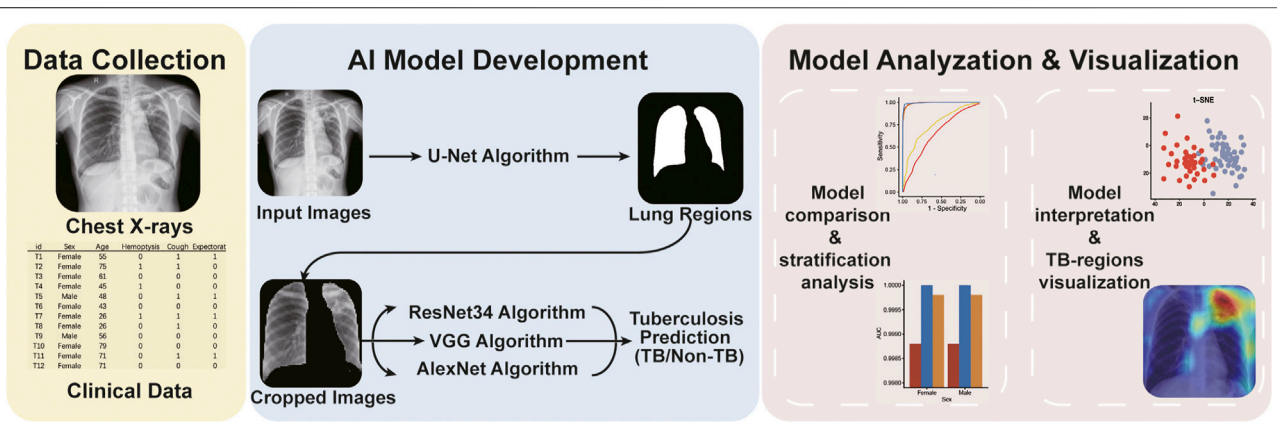


FIGURE 1 | The workflow of the study.

TABLE 1 | A summary of clinical characteristics of training and testing sets.

	Training set		Testing set	
	TB cases	**Non-TB cases**	**TB cases**	**Non-TB cases**
Sex				
Female	2,118	1885	463	478
Male	1882	1818	537	447
Age				
<65 years	2,297	1971	461	580
≥65 years	1703	1732	539	345
Symptoms				
Cough	3,304	6	782	8
Expectoration	2,726	4	606	1
Hemoptysis	461	5	128	4
Fever	1,563	4	429	0
Fatigue	1,143	0	248	0
Night Sweating	789	0	135	0
Bacteriological Test				
Sputum Culture/Smear Positive	807	0	297	0
Bacteriological Test Positive	1788	0	509	0
Bacteriological Test Negative	2,212	3,703	491	925
Total	4,000	3,703	1,000	925

Sputum Culture/Smear Positive: sputum culture positive or smear positive. Bacteriological Test Positive: sputum culture/smear positive or Xpert test positve. Bacteriological Test Negative: sputum culture negative, sputum smear negative and Xpert test negative.

(Krizhevsky et al., 2012). The structures of the three convolutional neural networks were illustrated in **Figure 2**. We loaded the pretrained model in ImageNet and replaced the last linear classification layer with a new linear layer for negative (individuals without TB) or positive (with TB) prediction. The ResNet model was trained for 120 epochs in total with chest X-ray images of the training set (**Supplementary Figure S1**), setting the initial learning rate at 1e−3 and using the inverse learning rate decay schedule. VGG and AlexNet were also trained with the same settings, and then three AI algorithms were validated on the testing set.

Comparison of AI Algorithms and Stratification Analysis

To evaluate the performance of the three models, the accuracy, sensitivity and specificity of the three models were calculated in both sets for comparison (**Table 2**), and the plots of receiver operating characteristic (ROC) curves with area under the curve (AUC) values were also generated with the pROC package. The AI algorithm with the best performance was chosen for the following analysis and visualization. To further investigate the reliability and robustness of the AI models, cases of the testing set were stratified into multiple subgroups based on their age, sex and respiratory symptoms, and then the AUC value of each model were calculated and compared within each subgroup with the pROC package.

Interpretability Analysis and Feature Visualization

After the comparison, ResNet algorithm was chosen for further interpretability analysis and feature visualization. After the global average pooling layer of ResNet, we obtained a 512-dimensional vector for each image. To visualize the learned feature, we used the t-SNE method, which could reduce the high-dimensional vector to a low-dimensional vector (Van et al., 2008). Here, we reduced each vector to 2-dimension and therefore, visualized the differences in high-dimensional complex features between the positive and negative images captured by the AI algorithm. We also obtained sets of feature maps from the last convolution layer of ResNet. Then, class activation maps (CAMs) were generated by the linear combination of the fully connected layer weights and feature maps (Zhou et al., 2016). Discriminative regions of the CAMs were in red (hot areas), indicating the AI-predicted TB regions.

RESULTS

In the beginning of the study, 9628 X-ray images with detailed clinical information were collected (**Figure 1**). Clinical characteristics of the collected cases were summarized in **Table 1**. Three different AI algorithms were trained on a large dataset containing 4,000 images from patients with TB and 3,703 images from individuals without TB (**Figure 2**). After training, these AI algorithms were compared to each other on the testing set and the training set. The accuracy, sensitivity and specificity of the three algorithms were all higher than 94% when they were tested on the testing set (**Table 2**). The ResNet model had the strongest diagnostic ability among the three AI algorithms, whose accuracy, sensitivity and specificity reached 96.73, 95.50 and 98.05%, respectively. The performance of the AI models were further investigated with ROC curves. All the AI algorithms performed well in both training set and testing set (**Figures 3A,B**), with AUC values higher than 0.99 (**Table 2**). The AUC value of the AlexNet, VGG and ResNet reached 0.9917, 0.9902 and 0.9944, respectively on the testing set. These results suggested that the ResNet algorithm outperformed the rest models, demonstrating its excellent diagnostic value for TB and was selected for further analysis and visualization.

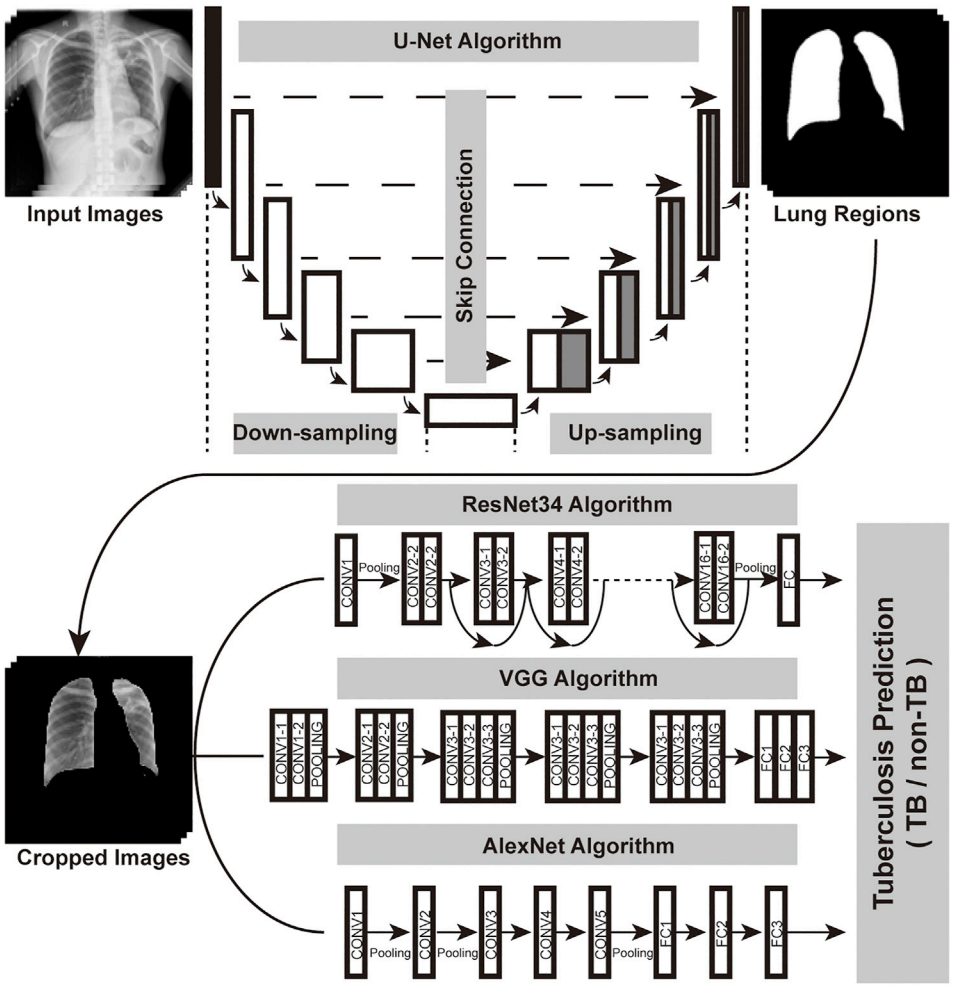


FIGURE 2 | Overall structure of the DCNN-based AI diagnosis system. The workflow of the system could be divided into two parts: image segmentation network (U-Net), image classification network (ResNet or VGG or AlexNet). Regions of the lung in the original chest X-ray photographs were recognized by the U-Net. Then, the cropped and resized lung region images served as an input for image classification algorithms, which generated diagnoses.

TABLE 2 | The performance of the models.

	Training set			AUC (95% CI)	Testing set			AUC (95% CI)
	Accuracy (%)	Sensitivity (%)	Specificity (%)		Accuracy (%)	Sensitivity (%)	Specificity (%)	
AlexNet	98.34	98.58	98.08	0.9988 (0.9984–0.9992)	95.06	93.20	97.08	0.9917 (0.9889–0.9945)
VGG	99.03	99.75	98.24	0.9998 (0.9997–0.9999)	94.96	94.20	95.78	0.9902 (0.9872–0.9932)
ResNet	99.92	99.90	99.95	1 (1–1)	96.73	95.50	98.05	0.9944 (0.9921–0.9967)

AUC, the area under the receiver operator characteristic curve; 95% CI: 95% confidence interval.

To understand how the AI algorithms distinguished TB radiographs from non-TB ones, the dimensional reduction t-SNE algorithm was applied to reduce the high-dimensional differences in visual-semantic information aggregated by the AI algorithm between TB-positive and TB-negative images into a two-dimensional plot. Here, taking ResNet as an example, an evident boundary was observed between the positive and negative image groups, while clustered distributions within each group were also observed, indicating that the ResNet algorithm succeeded in recognizing shared features of TB-positive images and features that distinguished the positive and negative images (**Figure 3C**).

Apart from evaluating the diagnostic accuracy on the entire testing set or training set, stratification analysis was also conducted to verified the reliability and the robustness of the

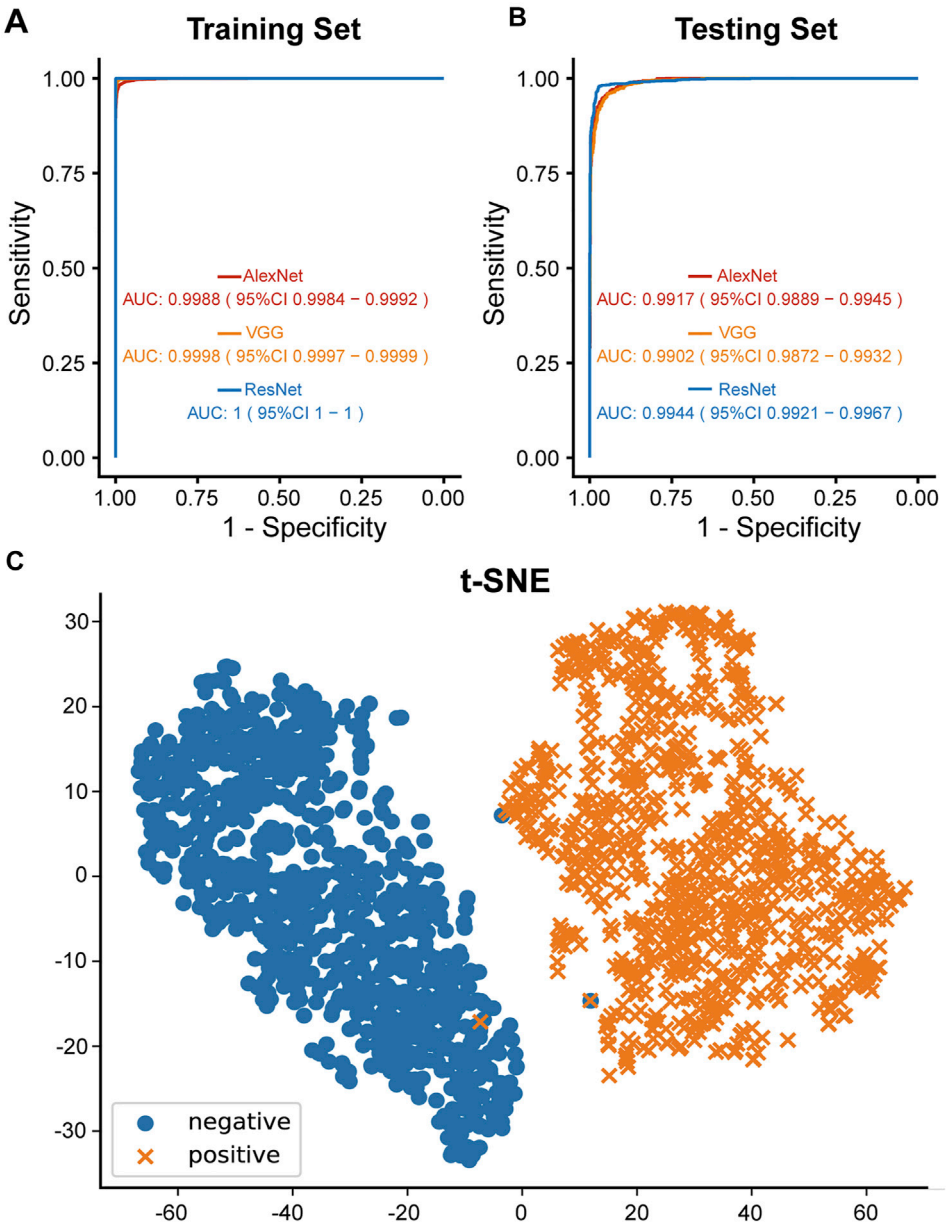


FIGURE 3 | Diagnostic ability of the AI models. **(A)** ROC curves of three different AI models of the training set. **(B)** ROC curves of three different AI models of the testing set. **(C)** Diagnostic ability of ResNet algorithm visualized by t-SNE algorithm. Blue and orange dots indicated TB and non-TB cases of the testing set, respectively.

ResNet algorithm. The ResNet model was the best performing model with highest AUC value in all subgroups of the training set and most subgroups of the testing set (**Figure 4**). The result of the stratification analysis suggested that the ResNet algorithm was capable of providing accurate TB diagnoses for people of all ages and both sexes or patients with different respiratory symptoms.

In order to step further in assisting the TB diagnosis, we visualized the ResNet algorithm recognized TB affected regions of the chest radiographs using the OpenCV package. Masked with red and alpha-blended with the black-and-white input X-ray image, "hot regions" drawn by AI provided accurate disease-affected areas and indicated high consistency with the TB regions mapped by experienced physicians and radiologists with bounding boxes (**Figure 5**). Our method not only provided a correct diagnosis of pulmonary TB but also identified precise TB regions with a heatmap, which has great potential in assisting the diagnosis of TB as an interpretable and reliable AI algorithm.

DISCUSSION

As an effective method for TB screening and diagnosis, chest radiography is recommended by multiple clinical guidelines

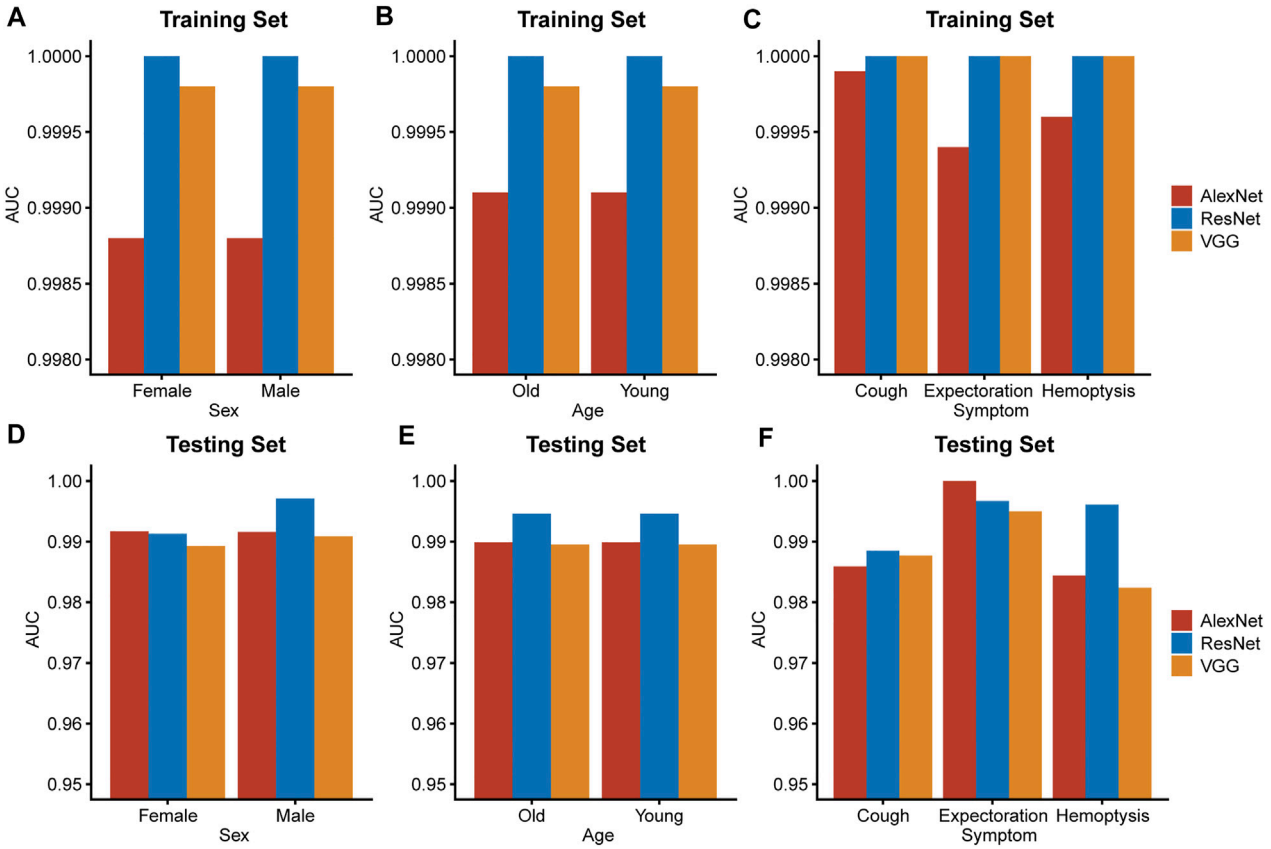


FIGURE 4 | Stratification analysis. Subgrouping by important clinical characteristics, including sex **(A,D)**, age **(B,E)** and respiratory symptoms **(C,F)**, AUC values of the three models were calculated and compared in both sets. Young: under 65 years old. Old: 65 years old or over 65 years old.

despite its high inter-radiologist and intra-radiologist variability, moderate specificity and other limitations. In contrast to manual diagnosis, deep learning based computer-assisted diagnostic systems have the potential to overcome the aforementioned drawbacks and provide professional diagnosis for TB. Here, we established a ResNet-based chest X-ray AI diagnosis system for TB, which provided accurate diagnoses and was capable of serving as triage tests at the bedside.

A previous study has reported that the AlexNet based machine learning algorithm can accurately classify TB based on chest radiography (Lakhani and Sundaram, 2017). Here, our results indicated that the ResNet-based AI was superior to AlexNet and VGG, which suggested that the ResNet diagnosis system would better assist physicians in diagnosing TB.

Low interpretability is one of the major inherent problems of the machine learning models, including deep learning algorithms. Due to the complex calculating process and tremendous amount parameters of the neural networks, it remains difficult for us to learn about how they work and why they come to certain conclusions that remarkably similar to human experts' opinions, suggesting that the neural networks, especially deep neural networks, are so-called black boxes (Wang et al., 2021). Even though depth of the DCNNs are becoming sheerer and sheer, many tools, including t-SNE and CAMs, are still available for the visualization of neural networks and breaking up the black boxes, convincing us that the neural network algorithm have the ability to recognized features of abnormality in the medical images rather than nonrelevant parts of the graphs. In this study, we wanted to know whether the trained DCNN based AI algorithm was focusing at regions of TB-associated abnormality in the lung. As illustrated in **Figure 5**, the discriminative regions, which were recognized by AI and masked with red, were the exact TB regions recognized and identified by doctors. This indicates that the ResNet-based AI algorithm not only provides doctors with highly accurate diagnoses but also interpretable marks of TB regions, which is of great help in analyzing chest X-ray images and recognizing TB in patients.

As a relatively remote and poor region with scarce medical resources in the past years, Xinjiang has been troubled by the continuous spread of TB (Zheng et al., 2021). Despite of financial difficulties, Xinjiang has managed to carry out many new policies and plans to boost investment in TB prevention and control, aiming at the early detection and

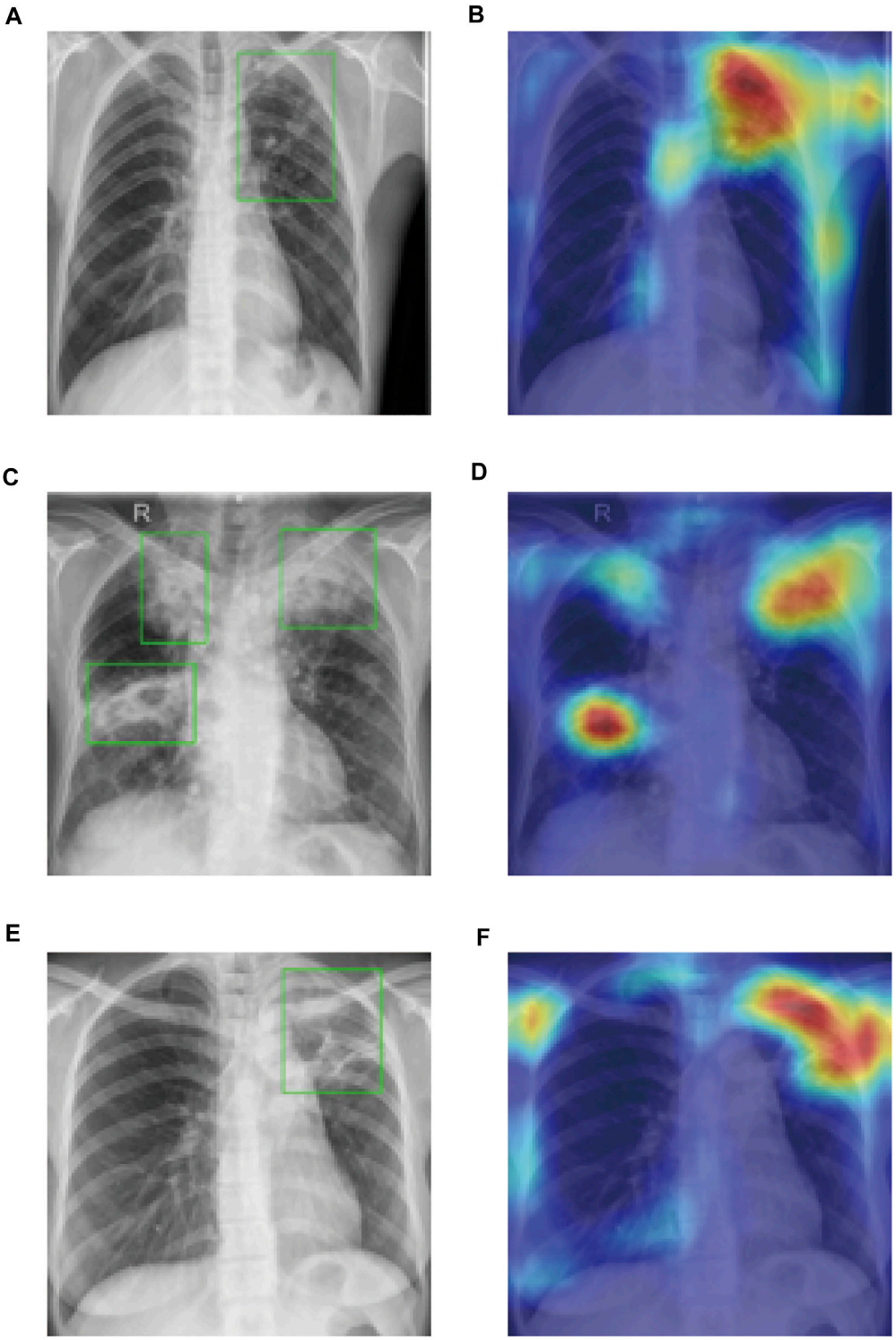


FIGURE 5 | CAMs generated by ResNet matched the precise regions of TB abnormalities. Bounding boxes [in **(A,C,E)**] meant the regions of abnormalities identified by doctors and hot regions [in **(B,D,F)**] showed the discriminative regions generated by AI algorithm.

proper treatment of TB cases. Early diagnosis of active TB is the key to controlling the rapid rise of TB incidence. Considering the excellent performance of the ResNet-based AI diagnosis system, it would greatly prompt the early diagnosis of active TB and help in preventing the spread of TB in Xinjiang.

However, there are several limitations in the study. First, although sputum culture is the gold standard for active pulmonary TB diagnosis, some patients of pulmonary TB have negative sputum culture results. Taking the results of multiple tests and typical clinical manifestation together, patients of TB were diagnosed, which remains risks of misdiagnoses and producing wrong labels for the chest radiographs. Besides, the chest radiographs were collected in two hospitals, suggesting that differences might exist between images captured in the two centers. In addition, we limited the study population to people aged ≥15 years, which also limited the generalizability of our AI diagnosis system towards pediatric cases.

In conclusion, our study established a ResNet-based AI diagnosis system that was effective in diagnosing active TB from chest radiographs without external clinical information assistance.

DATA AVAILABILITY STATEMENT

The clinical data and chest radiographs are not publicly available for patient privacy protection purposes. Any individual affiliated with an academic institution may request access to the original images and clinical data from the corresponding author (MN) for non-commercial, research purposes.

AUTHOR CONTRIBUTIONS

Study design and conception were proposed by GL and XZ; paper writing were done by MN, JM, CH, GL, and XZ; chest radiographs were collected by MN, AT, AA, AK, and XZ; AI model training, testing and visualizing were done by JM, CH, DZ, and GL. All authors interpreted the results and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the "Tianshan Innovation Team Plan" of the Autonomous Region (No. 2022D14007, MN), the Open Project of The State Key Laboratory for the Causes and Prevention of Middle Asian High Disease Jointly Built by the Province and the Ministry (No. SKL-HIDCA-2021-JH6, MN), Innovation Team Project of "Pearl River Scholar Tianshan Talents" Cooperative Expert Studio of The First People's Hospital of Kashi Prefecture (No. KDYY202017, MN).

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/fmolb.2022.874475/full#supplementary-material

REFERENCES

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, December 2016. doi:10.1109/cvpr.2016.90

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 60, 84–90. NIPS. doi:10.1145/3065386

Lakhani, P., and Sundaram, B. (2017). Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* 284 (2), 574–582. doi:10.1148/radiol.2017162326

Organization, W. H. (2016). *Chest Radiography in Tuberculosis Detection: Summary of Current WHO Recommendations and Guidance on Programmatic Approaches*. Geneva: World Health Organization.

Organization, W. H. (2021). *WHO Consolidated Guidelines on Tuberculosis: Module 2: Screening: Systematic Screening for Tuberculosis Disease*. Geneva: World Health Organization.

Qin, Z. Z., Ahmed, S., Sarker, M. S., Paul, K., Adel, A. S. S., Naheyan, T., et al. (2021a). Tuberculosis Detection from Chest X-Rays for Triaging in a High Tuberculosis-burden Setting: an Evaluation of Five Artificial Intelligence Algorithms. *The Lancet Digital Health* 3 (9), e543–e554. doi:10.1016/S2589-7500(21)00116-3

Qin, Z. Z., Naheyan, T., Ruhwald, M., Denkinger, C. M., Gelaw, S., Nash, M., et al. (2021b). A New Resource on Artificial Intelligence Powered Computer Automated Detection Software Products for Tuberculosis Programmes and Implementers. *Tuberculosis* 127, 102049. doi:10.1016/j.tube.2020.102049

Ragonnet, R., Trauer, J. M., Geard, N., Scott, N., and McBryde, E. S. (2019). Profiling *Mycobacterium tuberculosis* Transmission and the Resulting Disease burden in the Five Highest Tuberculosis burden Countries. *BMC Med.* 17 (1), 208. doi:10.1186/s12916-019-1452-0

Rajpurkar, P., O'Connell, C., Schechter, A., Asnani, N., Li, J., Kiani, A., et al. (2020). CheXaid: Deep Learning Assistance for Physician Diagnosis of Tuberculosis Using Chest X-Rays in Patients with HIV. *Npj Digit. Med.* 3, 115. doi:10.1038/s41746-020-00322-2

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Editors N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (New York City, NY: Springer International Publishing), 234–241.

Simonyan, K., and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. Available: https://ui.adsabs.harvard.edu/abs/2014arXiv1409.1556S (Accessed September 01, 2014).

Tusun, D., Abulimiti, M., Abulimiti, M., Mamuti, X., Liu, Z., Xu, D., et al. (2021). The Epidemiological Characteristics of Pulmonary Tuberculosis - Kashgar Prefecture, Xinjiang Uygur Autonomous Region, China, 2011-2020. *China CDC Wkly* 3 (26), 557–561. doi:10.46234/ccdcw2021.144

Van, D. M. L., and Hinton, G. (2008). Visualizing High-Dimensional Data Using T-SNE. *J. Machine Learn. Res.* 9 (2), 2579–2605.

Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., et al. (2021). CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Trans. Vis. Comput. Graphics* 27 (2), 1396–1406. doi:10.1109/TVCG.2020.3030418

Zheng, Y., Zhang, X., Wang, X., Wang, K., and Cui, Y. (2021). Predictive Study of Tuberculosis Incidence by Time Series Method and Elman Neural

Network in Kashgar, China. *BMJ Open* 11 (1), e041040. doi:10.1136/bmjopen-2020-041040

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning Deep Features for Discriminative Localization," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016. doi:10.1109/cvpr.2016.319

Zunair, H., and Ben Hamza, A. (2021). Sharp U-Net: Depthwise Convolutional Network for Biomedical Image Segmentation. *Comput. Biol. Med.* 136, 104699. doi:10.1016/j.compbiomed.2021.104699

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 Nijiati, Ma, Hu, Tuersun, Abulizi, Kelimu, Zhang, Li and Zou. 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 Genetics

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fgene.2022.872238

DNA Repair–Related Gene Signature in Predicting Prognosis of Colorectal Cancer Patients

Min-Yi Lv[1,2†], Wei Wang[3†], Min-Er Zhong[1,2], Du Cai[1,2], Dejun Fan[1,2,4], Cheng-Hang Li[1,2], Wei-Bin Kou[5], Ze-Ping Huang[1,2], Xin Duan[1,2], Chuling Hu[1,2], Qiqi Zhu[1,2], Xiaosheng He[1,2] and Feng Gao[1,2]**

[1]Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2]Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [3]Biomedical Big Data Center, Huzhou Maternity & Child Health Care Hospital, Huzhou, China, [4]Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Supported by National Key Clinical Discipline, Guangdong Institute of Gastroenterology, Guangzhou, China, [5]The University of Hong Kong, Hong Kong, Hong Kong SAR, China

OPEN ACCESS

Edited by:
Ka-Chun Wong,
City University of Hong Kong, Hong Kong SAR, China

Reviewed by:
Chieh-Lin Jerry Teng,
Taichung Veterans General Hospital, Taiwan
Petter Fruhling,
Uppsala University, Sweden

***Correspondence:**
Xiaosheng He
hexsheng@mail.sysu.edu.cn
Feng Gao
gaof57@mail.sysu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 09 February 2022
Accepted: 07 March 2022
Published: 11 April 2022

Citation:
Lv M-Y, Wang W, Zhong M-E, Cai D, Fan D, Li C-H, Kou W-B, Huang Z-P, Duan X, Hu C, Zhu Q, He X and Gao F (2022) DNA Repair–Related Gene Signature in Predicting Prognosis of Colorectal Cancer Patients. Front. Genet. 13:872238. doi: 10.3389/fgene.2022.872238

Background: Increasing evidence have depicted that DNA repair–related genes (DRGs) are associated with the prognosis of colorectal cancer (CRC) patients. Thus, the aim of this study was to evaluate the impact of DNA repair–related gene signature (DRGS) in predicting the prognosis of CRC patients.

Method: In this study, we retrospectively analyzed the gene expression profiles from six CRC cohorts. A total of 1,768 CRC patients with complete prognostic information were divided into the training cohort (n = 566) and two validation cohorts (n = 624 and 578, respectively). The LASSO Cox model was applied to construct a prediction model. To further validate the clinical significance of the model, we also validated the model with Genomics of Drug Sensitivity in Cancer (GDSC) and an advanced clear cell renal cell carcinoma (ccRCC) immunotherapy data set.

Results: We constructed a prognostic DRGS consisting of 11 different genes to stratify patients into high- and low-risk groups. Patients in the high-risk groups had significantly worse disease-free survival (DFS) than those in the low-risk groups in all cohorts [training cohort: hazard ratio (HR) = 2.40, $p < 0.001$, 95% confidence interval (CI) = 1.67–3.44; validation-1: HR = 2.20, $p < 0.001$, 95% CI = 1.38–3.49 and validation-2 cohort: HR = 2.12, $p < 0.001$, 95% CI = 1.40–3.21). By validating the model with GDSC, we could see that among the chemotherapeutic drugs such as oxaliplatin, 5-fluorouracil, and irinotecan, the IC50 of the cell line in the low-risk group was lower. By validating the model with the ccRCC immunotherapy data set, we can clearly see that the overall survival (OS) of the objective response rate (ORR) with complete response (CR) and partial response (PR) in the low-risk group was the best.

Abbreviations: DRGs, DNA repair-related genes; DRGS, DNA repair-related gene signature; CRC, colorectal cancer; DEGs, differentially expressed genes; HR, hazard ratio; GEO, Gene Expression Omnibus; GSEA, Gene Set Enrichment Analysis; LASSO, least absolute shrinkage and selection operator; OS, overall survival; ROC, receiver operating characteristic; and TCGA, The Cancer Genome Atlas.

Conclusions: DRGS is a favorable prediction model for patients with CRC, and our model can predict the response of cell lines to chemotherapeutic agents and potentially predict the response of patients to immunotherapy.

Keywords: DNA repair–related genes, prognostic, colorectal cancer, immunotherapy, microsatellite instability

BACKGROUND

With the third highest incidence rate in the world, colorectal cancer (CRC) is a serious threat to human health (Bray et al., 2018). Nowadays, due to lifestyle changes, there is an increasingly high incidence of mortality from CRC (Zheng et al., 2014). As one of the most common gastrointestinal tumors in general surgery, CRC is a multifactorial disease with extremely complex pathogenesis (Migliore et al., 2011). At present, the early diagnosis of CRC has involved epigenetics, genomics, and so on (Marcuello et al., 2019). DNA repair is a series of processes by which a cell recognizes and corrects damage to the DNA molecules that encode its genome (Zinovkina, 2018; Burdak-Rothkamm and Rothkamm, 2021), and it is extremely important for maintaining the stability of the genome and protecting the genome from damage by endogenous and environmental agents (Friedberg, 2001). It is estimated that human cells suffer more than 2×10^4 DNA damage events per day (Lindahl and Wood, 1999), but generally speaking, cells can respond to this damage through efficient and highly regulated DNA repair mechanisms (Lindahl and Wood, 1999; Iyama and Wilson, 2013). Repair mechanisms include nuclear excision repair, base excision repair, mismatch repair (MMR), and double-strand break repair (Iyama and Wilson, 2013). As we all know, genomic instability caused by the destruction of DNA damage and repair mechanism can lead to cancer progression, and DNA repair genes are often found to mutate in cancer (Knijnenburg et al., 2018). Recently, Knijnenburg et al. (2018) discovered mutations related to DNA damage response genes by analyzing the TCGA data and found that several mutations in DNA damage response and repair genes occur in the colon adenocarcinoma and rectal adenocarcinoma data sets.

Due to the limited options for capturing the molecular heterogeneity of the disease and the lack of consideration and sufficient validation of other gene expressions, few of the prognostic models of early stage CRC have been applied in clinical practice (Guinney et al., 2015; Phipps et al., 2015). Thus, an accurate method is needed to identify effective prognostic models to assess the disease-free survival (DFS) of patients with CRC. The aim of the present study is to examine the interrelationships between DNA repair–related genes (DRGs) and CRC, to determine an effective prognostic model to evaluate the DFS of patients with CRC and provide guidance for clinicians in early diagnosis and treatment.

MATERIALS AND METHODS

Patients

We retrospectively analyzed the gene expression profiles of CRC samples from six public cohorts. Totally, 1,768 samples were available for analysis in the current study. The CIT/GSE39582 (n = 566) was used for training the model, and The Cancer Genome Atlas colorectal cancer (TCGA, n = 624) was selected to serve as a validation-1 cohort. The remaining four microarray data sets (GSE14333, GSE33113, GSE37892, and GSE39084) were merged into a validation-2 cohort (n = 578) (**Table 1**). The transcriptome RNA-sequencing data of the CRC samples were from the TCGA data portal, and other microarray data sets were acquired directly from the GEO database. The institutional review board of our hospital approved this study, and data were collected from 12 May to 10 October 2020.

Construction and Validation of DNA Repair–Related Gene Signature

Firstly, a complete list of DRGs was available online from MSigDB (version 6.2, https://www.gsea-msigdb.org/gsea/msigdb). We identified a list of candidate genes differentially expressed between relapsed samples and non-relapsed samples by using the "limma" R package (Diboun et al., 2006). The genes with an absolute log2-fold change of more than 1 and an adjusted $p < 0.05$ were considered for subsequent analysis. In order to minimize over-fitting risk, we applied a Cox proportional hazards regression model on CRC samples combined with the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1997). The penalty parameter was estimated by 10-fold cross-validation in the training data set at the minimum partial likelihood deviance.

We divided patients into high-risk and low-risk groups by determining the optimal threshold through the time-dependent receiver operating characteristic (ROC) curve (survivalroc, version 1.0.3) at 5 y in the training data set. The ROC curve was estimated by the Kaplan–Meier estimation method. We performed univariate and multivariate Cox regression analyses of the cohort to verify that the 11-DRG signature was independent of other clinical features.

Functional Annotation Analysis

To evaluate the biological functions of the DNA repair–related gene signature (DRGS), enrichment analysis for differentially expressed genes in different groups was applied using the R package "gProfileR." We used the Bioconductor package "HTSanalyzeR" to perform Gene Set Enrichment Analysis (GSEA) to predict significant dysregulated pathways (Subramanian et al., 2005; Wang et al., 2011). Gene sets of cancer hallmarks from MSigDB (Liberzon et al., 2015) were examined.

Validation of Genomics of Drug Sensitivity in Cancer Database, Immunotherapy Database, and Tumor Immune Dysfunction and Exclusion

To further explore the clinical application of our model, we used Genomics of Drug Sensitivity in Cancer (GDSC) to analyze the

TABLE 1 | Characteristics of cohorts included in this study.

Characteristics	Training cohort GSE39582	Validation-1 TCGA	Validation-2 (combination of GSE14333, GSE33113, GSE37892, and GSE39084)
Number of patients	566	624	578
Mean age	66.85	66.27	66.37
Gender			
Male	256 (45.23%)	292 (46.79%)	270 (46.71%)
Female	310 (54.77%)	332 (53.21%)	308 (53.29%)
TNM stage			
Stage I	37 (6.54%)	105 (16.83%)	53 (9.17%)
Stage II	264 (46.64%)	230 (36.86%)	280 (48.44%)
Stage III	205 (6.22%)	180 (28.85%)	164 (28.37%)
Stage IV	60 (10.60%)	88 (14.10%)	81 (14.01%)
NA	-	21 (3.37%)	-
Tumor location			
Left	342 (60.42%)	354 (56.73%)	269 (46.54%)
Right	224 (39.58%)	270 (43.27%)	216 (37.73%)
NA	-	-	93 (16.09%)
RFS event			
Yes	177 (30.62%)	100 (16.03%)	130 (22.50%)
No	380 (65.74%)	416 (66.67%)	382 (66.09%)
NA	9 (1.56%)	108 (17.31%)	66 (11.42%)
MMR status			
MSI	75 (13.25%)	189 (30.29%)	44 (7.61%)
MSS	444 (78.45%)	431 (69.07%)	114 (19.72%)
NA	47 (8.30%)	4 (0.64%)	420 (72.66%)
CIMP status			
Positive	91 (16.07%)	NA	39 (6.75%)
Negative	405 (71.56%)	NA	118 (20.42%)
NA	70 (12.37%)	624 (100%)	421 (72.84%)
TP53 status			
Wide-type	161 (28.45%)	-	39 (6.75%)
Mutation	190 (33.57%)	-	29 (5.02%)
NA	215 (37.99%)	-	510 (88.24%)
KRAS status			
Wide-type	328 (57.95%)	34 (5.45%)	110 (19.03%)
Mutation	217 (38.34%)	30 (4.81%)	48 (8.30%)
NA	21 (3.71%)	560 (89.74%)	420 (72.66%)
BRAF status			
Wide-type	461 (81.45%)	32 (5.13%)	133 (23.01%)
Mutation	51 (9.01%)	3 (0.48%)	25 (4.33%)
NA	54 (9.54%)	589 (94.39%)	420 (72.66%)

differences of chemotherapeutic drugs between the high-risk group and the low-risk group.

As known, immunotherapy is a hot topic, and we want to know whether this model can predict immunotherapy. We verified our model by using the data provided in the article "Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma (ccRCC)" published in *Nature Medicine* (Braun et al., 2020). We constructed the DRGS in the data set of advanced clear cell renal cell carcinoma and divided it into the high-risk and low-risk groups according to the cutoff of our original model. The overall survival (OS) curve was drawn using the Kaplan–Meier method. In addition, we selected some immune-related indicators in the data set and compared the differences of these indicators between the high- and low-risk groups by *t*-test. Besides, we also analyzed the OS curve of the objective response rate (ORR) of immunotherapy.

The tumor immune dysfunction and exclusion (TIDE) algorithm can be used to predict the tumor response to immune checkpoint inhibition treatment and the function of genes regulating tumor immunity, so as to effectively predict the effect of immune checkpoint inhibition treatment.

Statistical Analysis

All the statistical analyses were performed on R (version 3.4.3, www.r-project.org). The hazard ratios were calculated using the "survcomp" package28 (version: 1.28.4) (Schröder et al., 2011). The LASSO regression was implemented using "glmnet" R package (version: 2.0.16). Cox regression analysis was used for single-factor and multifactor analyses of the results, and the receiver operating characteristic (ROC) curve and C-index were used to evaluate the model. A *p*-value of less than 0.05 was defined as statistical significance in all tests.

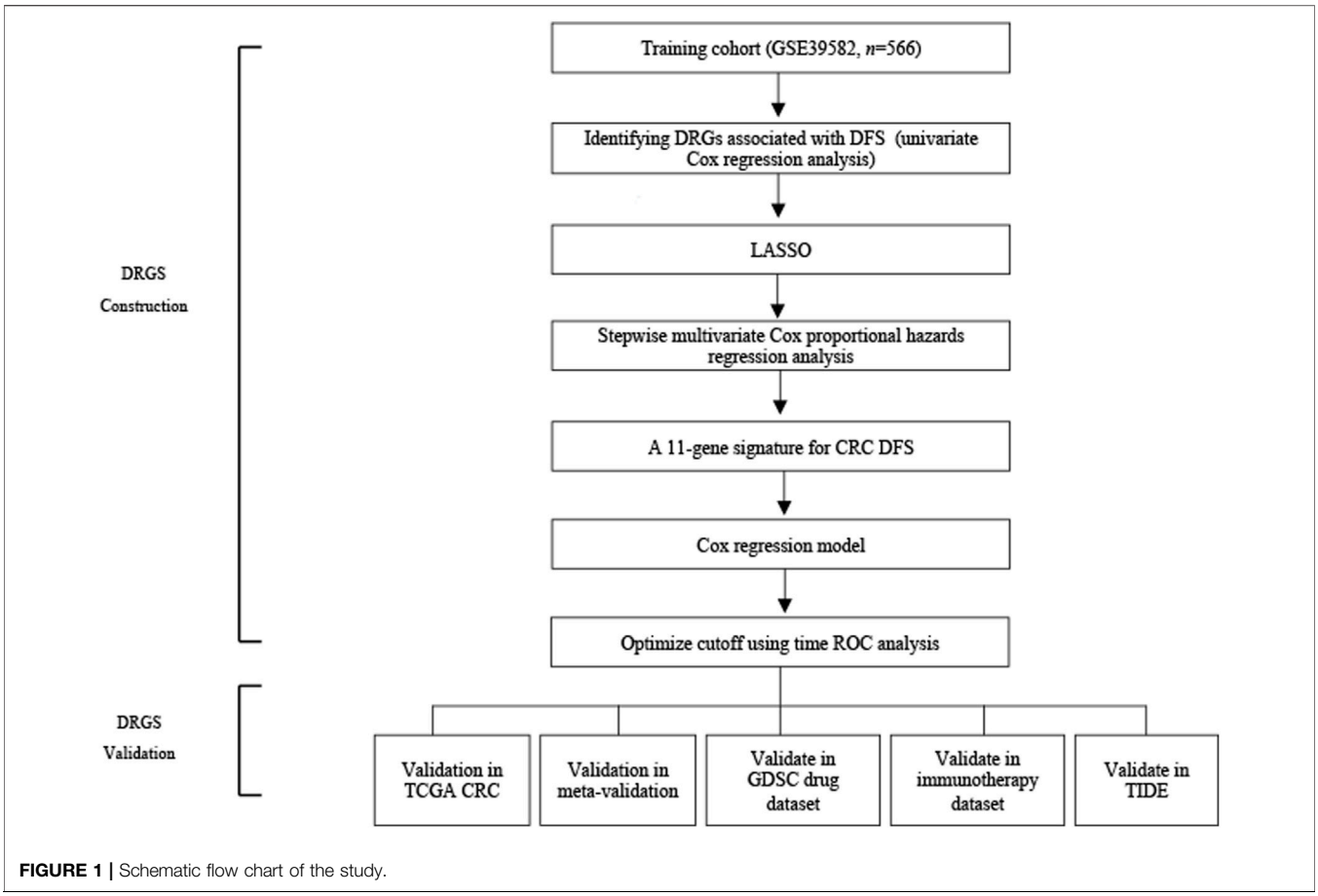


FIGURE 1 | Schematic flow chart of the study.

RESULTS

Construction and Definition of the DNA Repair-Related Gene Signature

A total of 1,768 CRC patients were included in the analysis. The CIT data set (GSE39582, $n = 566$) was used as the training cohort and genes with relatively high variation were maintained as candidates (**Table 1**, **Figure 1**). With median absolute deviation >0.5 and excluding the genes expressed less in the median expression level, 1,286 genes were screened out of 1,376 DRGs measured on all platforms from the data sets. In addition, in order to improve the robustness of the identification for the limited sample size, we further selected DRGs by using the Cox proportional hazards regression against 1,000 randomized trials (80% portion of samples each time) to assess the correlation between each candidate gene and patients' DFS in the training cohort. A total of 46 DRGs were robustly associated with individual patients' DFS. In order to minimize the over-fitting risk, we applied a Cox proportional hazards regression model to the CRC samples combined with the LASSO. By using LASSO Cox regression, 11 prognostic DRGs were selected and combined for the construction of DRGS (**Figures 2A,B**). The risk scores were calculated by the formula designed by the Cox regression model. The total risk score was imputed as follows (−0.1145 × POLR2B) + (−0.0653 × RAD1) + (0.0370 × CDA) + (0.1711 × NPR2) + (−0.0328 × UBE2D2) + (−0.0992 × BCL2) + (−0.0473 × PLD6) + (0.0896 × ERBB2) + (0.1220 ×ARPC1B) + (−0.1086 × FUT4) + (−0.0765 × PSME2). The time-dependent ROC curve analysis showed that the optimal cutoff to stratify high- and low-risk groups was −0.147 (**Figure 2C**).

Prognostic Evaluation of the DNA Repair-Related Gene Signature

Six colorectal cancer transcription data sets containing prognostic data were selected to assess the prognostic ability of the DRGS. The GSE39582 data set ($n = 566$) was used as a training data set (**Figure 2D**). The TCGA CRC dataset was enrolled as validation-1 cohort ($n = 624$), and additional data sets from the GEO were combined as validation-2 cohort ($n = 578$). Among the patients in the training and validation cohorts, more recurrences were found in the high-risk group than in the low-risk group (**Figures 3A,D,G**). When applied to a follow-up duration, the promising prognostic values of 2-, 3-, and 5-year AUC were 0.640, 0.664, and 0.653, respectively, in the training cohort. In the validation-1 cohort, the values of 2-, 3-, and 5-year AUC were 0.620,

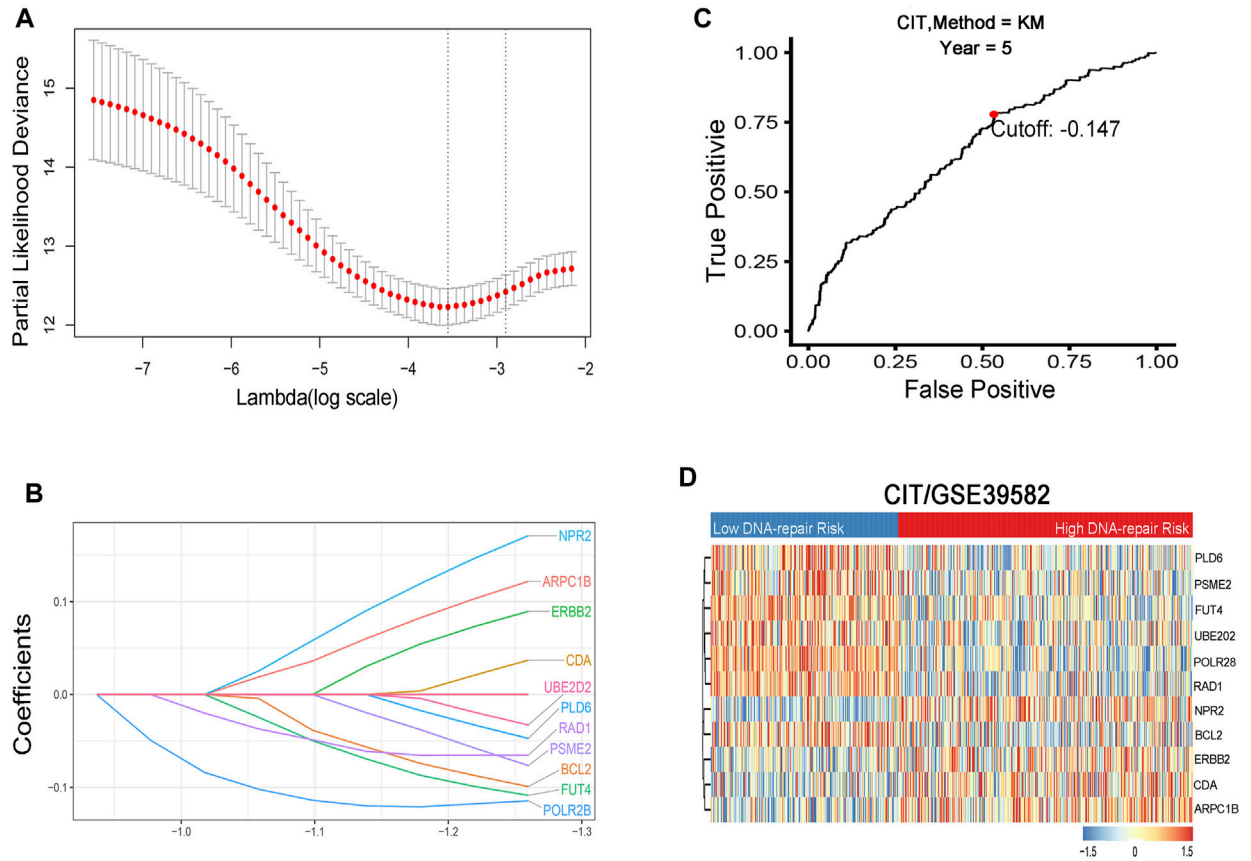


FIGURE 2 | (A) Identification and selection of prognostic genes by LASSO Cox proportional hazards regression. **(B)** Establishment of 11 DNA repair–related genes signature from the LASSO COX regression. **(C)** Optimal cutoff point of the prognostic gene signature at 5-y OS endpoint from the ROC curve. **(D)** Heat map of the 11 DNA repair–related genes in two risk groups.

0.628, and 0.606, respectively. Furthermore, in the validation-2 cohort, the values of 2-, 3-, and 5-year AUC were 0.645, 0.631, and 0.638, respectively (**Figures 3B,E,H**). The DRGS significantly stratified patients into the high- and low-risk groups in the training cohort (HR = 2.40, 95% CI = 1.67–3.44, $p < 0.001$), validation-1 cohort (HR = 2.20, $p < 0.001$, 95% CI = 1.38–3.49), and validation-2 cohort (HR = 2.12, $p < 0.001$, 95% CI = 1.40–3.21) (**Figures 3C,F,I**). Besides, the OS in the low-risk group was better than in the high-risk group (**Supplementary Figure 1**).

Compared to the risk scores calculated using the FDA-approved assay Oncotype DX colon algorithm, we found that the DRGS achieved better survival correlation in the training cohort (C-index, 0.78 vs 0.60), validation-1 cohort (C-index, 0.65 vs 0.51), and validation-2 cohort (C-index, 0.66 vs 0.62) (**Table 2**).

To further investigate whether the DRGS could serve as an independent predictor of prognosis, univariate and multivariate Cox proportional hazards regression analyses were performed. As expected, age, sex, tumor stage, tumor location, and pathologic gene status were associated with outcomes for CRC patients (**Table 3**). In the univariate analysis, DRGS, MMR status, and KRAS mutation status were significantly correlated with worse prognosis in the training cohort. After adjusting for clinical features such as age, gender, tumor location, and molecular types, the DRGS remained an independent prognostic factor in the multivariate analyses in both validation cohorts.

Functional Annotation of Genomics of Drug Sensitivity in Cancer

Gene Ontology (GO) analysis revealed that some biological process pathways (extracellular region, cell proliferation, and cell adhesion) were the main enriched pathways in the high-risk group (**Figure 4A**). In addition, the GSEA in the high-risk group when compared with the low-risk groups shown that the metastasis-related pathways (i.e., angiogenesis, KRAS signaling, epithelial mesenchymal transit, and myogenesis pathways) were enriched in the high-risk group (**Figure 4B**, **Supplementary Table S1**). Similarly, we obtained consistent results in the TCGA and validation-2 cohorts (**Supplementary Figure 2**). These findings suggest that the enrichment of pathways provided evidence of molecular mechanisms affected by the DRGS and thus can predict the prognosis of CRC.

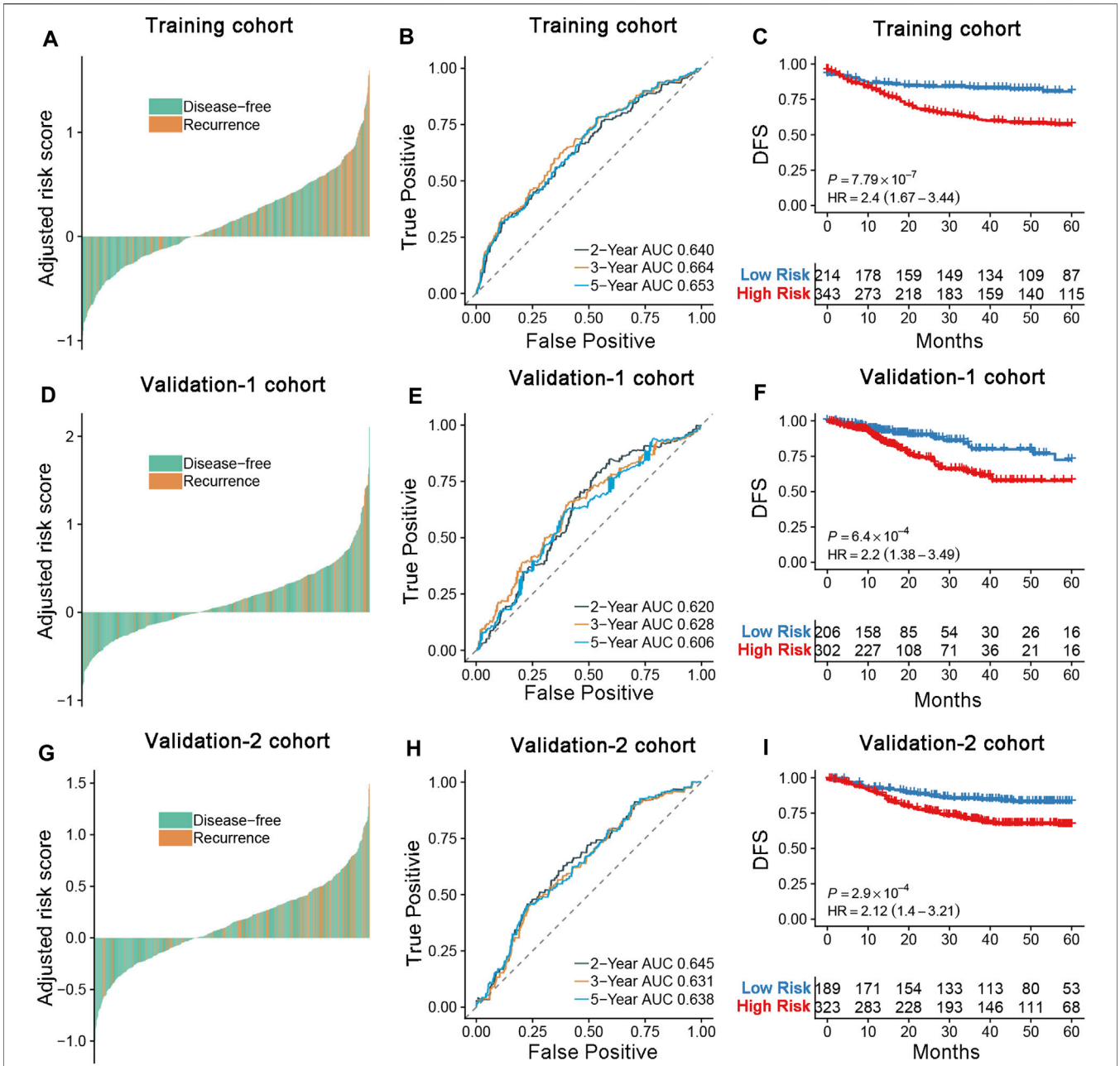


FIGURE 3 | (A,D,G) Distribution of the DRGS risk score and its correlation to recurrence in the training, validation-1, and validation-2 cohort. **(B,E,H)** Time-dependent ROC analysis of disease-free survival for CRC patients in the training, validation-1, and validation-2 cohorts at the time points of 2, 3, and 5 y. **(C,F,I)** Kaplan–Meier curves comparing survival of patients within the low- and high-risk groups in the training cohort, validation-1, and validation-2 cohorts. *p*-values were calculated using log-rank tests.

TABLE 2 | C-index for DRGS risk compared with Oncotype DX.

Cohorts	DNA repair risk		Oncotype DX colon	
	C-index	95% CI	C-index	95% CI
Training cohort	0.78	0.69–0.86	0.60	0.52–0.68
Validation-1 cohort	0.65	0.51–0.79	0.51	0.37–0.65
Validation-2 cohort	0.66	0.55–0.76	0.62	0.53–0.70

Validation of Genomics of Drug Sensitivity in Cancer Database and Immunotherapy Database

As known, MSI/MMR-deficient (dMMR) is widely considered as a promising biomarker, suggesting greater efficacy for immune checkpoint inhibitor (ICB) (Zhao et al., 2019). In order to further investigate the clinical application of our

TABLE 3 | Univariate and multivariate analyses of DRGS, and clinical and pathologic factors.

Characteristic	Training cohort GSE39582				Validation-1 TCGA				Validation-2 (GSE14333,GSE33113,GSE37892, and GSE39084)			
	Univariate		Multivariate		Univariate		Multivariate		Univariate		Multivariate	
	HR (95%CI)	*P*	HR (95%CI)	*P*	HR (95%CI)	*P*	HR (95%CI)	*P*	HR (95%CI)	*P*	HR (95%CI)	*P*
DRGS	2.40 (1.67–3.44)	9.40E-07	1.80 (1.22–2.64)	0.0028	2.20 (1.38–3.49)	<0.001	1.85 (1.13–3.02)	0.015	2.12 (1.40–3.21)	<0.001	1.75 (1.15–2.65)	0.0085
Gender	1.21 (0.89–1.64)	0.21	-	-	1.25 (0.82–1.89)	0.3	-	-	1.06 (0.74–1.50)	0.77	-	-
Age	1.00 (0.99–1.01)	1	-	-	0.99 (0.97–1.00)	0.16	-	-	0.99 (0.98–1.00)	0.16	-	-
Tumor location	1.27 (0.92–1.74)	0.14	-	-	0.88 (0.58–1.33)	0.55	-	-	1.19 (0.81–1.75)	0.38	-	-
MMR status	2.76 (1.46–5.24)	0.0012	1.94 (0.97–3.87)	0.062	0.94 (0.61–1.45)	0.77	-	-	1.74 (0.80–3.77)	0.15	-	-
CIMP status	0.71 (0.44–1.15)	0.16	-	-	-	-	-	-	0.82 (0.39–1.72)	0.6	-	-
CIN status	1.09 (0.70–1.71)	0.71	-	-	-	-	-	-	-	-	-	-
TP53 mutation	1.41 (0.99–2.03)	0.059	-	-	-	-	-	-	1.59 (0.70–3.61)	0.26	-	-
KRAS mutation	1.40 (1.03–1.91)	0.031	1.16 (0.84–1.59)	0.37	0.76 (0.31–1.83)	0.54	-	-	1.25 (0.65–2.38)	0.5	-	-
BRAF mutation	0.96 (0.54–1.69)	0.88	NA	-	0.00 (0.00–Inf)	0.46	-	-	1.61 (0.77–3.37)	0.2	-	-

model, we used GDSC to analyze the differences of chemotherapeutic drugs between the high-risk and low-risk groups. We selected 48 kinds of cell lines related to CRC. After dividing the cell lines into the high-risk and low-risk groups according to the cutoff of our model, we selected the chemotherapeutic drugs commonly used in clinics to see the IC50 of the cell lines in the high-risk and low-risk groups. We can see that among the chemotherapeutic drugs such as oxaliplatin, 5-fluorouracil, and irinotecan, the IC50 of the cell line in the low-risk group was lower (**Figure 5**). It showed that the cell lines in the low-risk group were more sensitive to these three drugs.

To examine whether the DRGS could predict the survival for ccRCC patients, the patients were divided into the high-risk and the low-risk groups according to the cutoff of our original model. The cutoff was still −0.147, and the prognosis data of these patients were analyzed. The OS of the high-risk group was worse than that of the low-risk group in ccRCC patients (HR = 1.45, 95% CI = 1.09–1.92, $p = 0.0103$) (**Figure 6A**). When it comes to the ORR of immunotherapy, we can clearly see the ORR with complete response (CR) and partial response (PR) that had better OS for both the high-risk and low-risk groups ($p < 0.001$). Notably, the OS of the low-risk group with the CR + PR was the best (**Figure 6B**).

Validation of Tumor Immune Dysfunction and Exclusion Database

We applied the TIDE algorithm which can predict the response to immunotherapy. The low-risk group had a lower TIDE score in GSE39582 and TCGA data sets, indicating that this subgroup was most likely to benefit from immunotherapy. Besides, the low-risk group had higher interferon gamma (IFNG), higher microsatellite instability (MSI) score, and lower cancer-associated fibroblasts (CAFs) amount, which confirmed the more activated immune landscape in this subgroup (**Figure 7**).

DISCUSSION

Colorectal cancer is the leading cause of death among gastrointestinal cancers. The incidence and mortality from colorectal cancer are increasing year by year, and its prognosis is closely related to early diagnosis (Siegel et al., 2016; Siegel et al., 2017). Numerous studies have highlighted the biomarkers that are associated with the pathogenesis and biology of CRC (Shah et al., 2014; De Rosa et al., 2016; Lech et al., 2016; Das et al., 2017), and many multigene prognostic signatures have been developed for CRC (Shah et al., 2014; Kandimalla et al., 2018; Ozawa et al., 2018; Gao et al., 2019; Kandimalla et al., 2019). Unfortunately, the accuracy of their prognosis predictions remains uncertain (Fung et al., 2014). We still need much more effort to achieve good prognostic CRC prediction, which is still considered a challenge.

In recent years, some studies have found some new results in DNA pathway repair and DRGs research. DRGs inactivation may

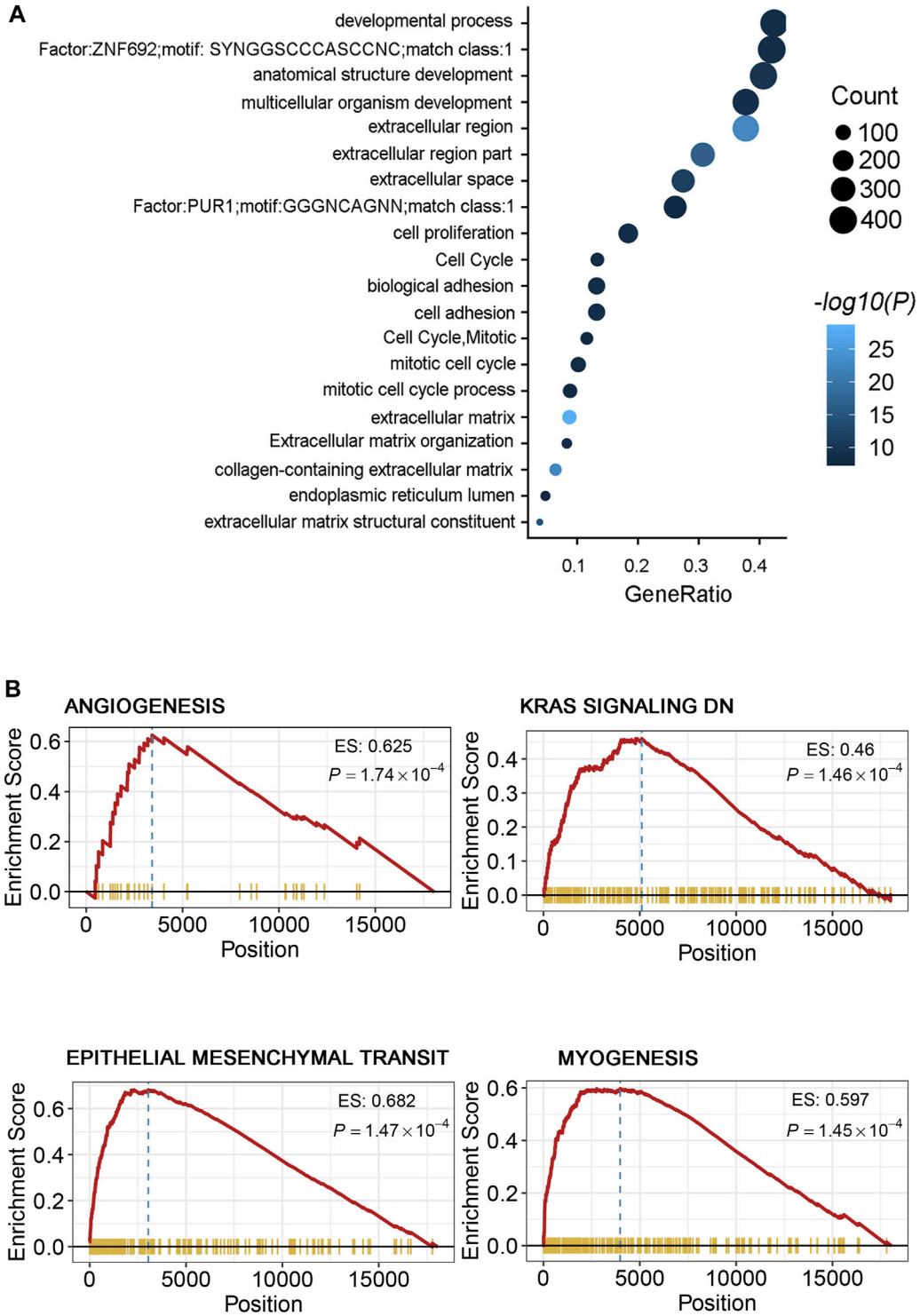


FIGURE 4 | (A) Gene ontology of the differentially expressed genes between the two risk groups. "GeneRatio" is the percentage of total differential genes in the given GO term. **(B)** GSEA showed several metastasis-related processes enriched in the high-risk group, including angiogenesis, KRAS signaling, epithelial mesenchymal transit (EMT), and myogenesis signal pathways.

disrupt genomic integrity, which may increase the risk of accumulation of gene mutations associated with cancer development (Bouwman and Jonkers, 2012). MSI/dMMR is widely considered as a promising biomarker, suggesting greater efficacy for ICB despite some limitations (Zhao et al., 2019). In this study, our purpose was to identify and validate a

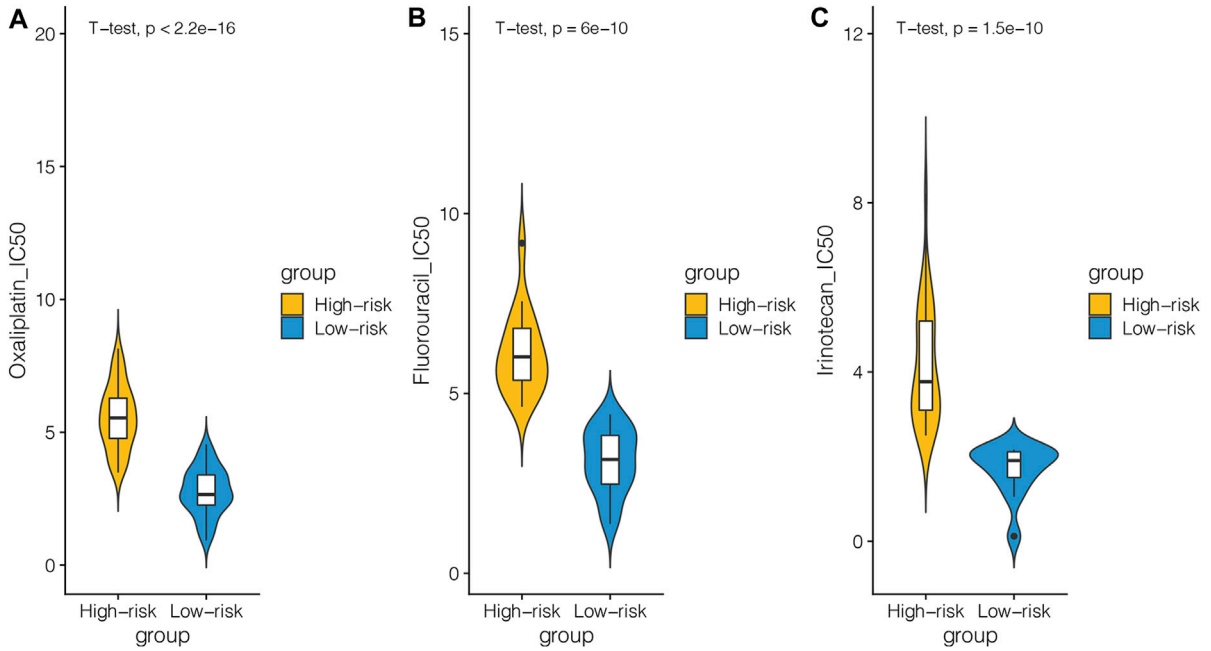


FIGURE 5 | CRC cell lines in the GDSC database were divided into the high-risk and low-risk groups based on DNA repair–related signature and the differences in response to chemotherapies between the two groups were analyzed. **(A)** Relationship between the cell line of the high-risk and low-risk groups and IC50 of oxaliplatin. **(B)** Relationship between the cell line of the high-risk and low-risk groups and IC50 of fluorouracil. **(C)** Relationship between the cell line of the high-risk and low-risk groups and IC50 of irinotecan.

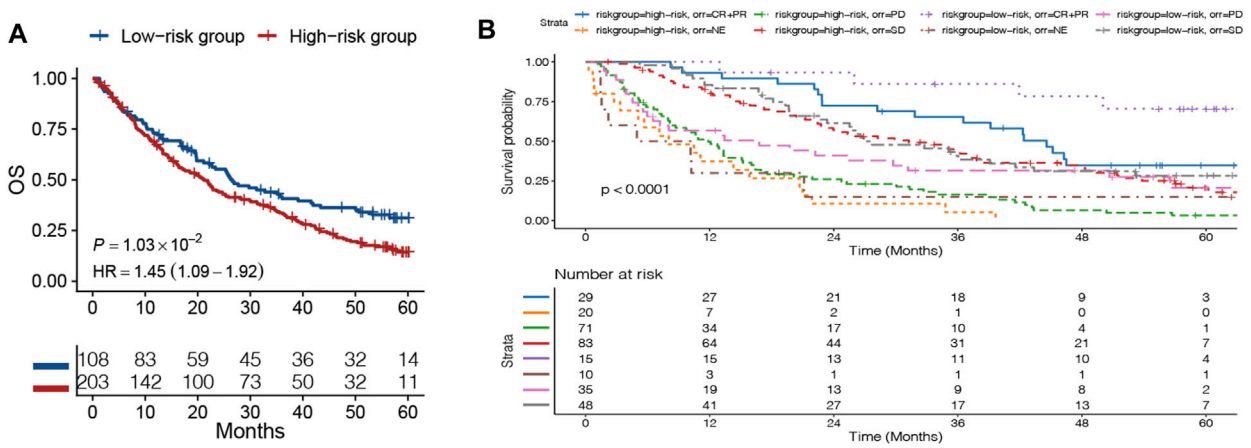


FIGURE 6 | Patients in the advanced clear cell renal cell carcinoma (ccRCC) database were divided into the high-risk and low-risk groups based on the DNA repair–related signature. **(A)** Kaplan–Meier curves comparing the survival of patients within the low- and high-risk groups in the ccRCC database. **(B)** OS curve of the objective response rate (ORR) of immunotherapy in ccRCC database.

reliable DRGS and improve the accuracy of survival prediction for CRC patients.

A total of 1,768 CRC patients from one training cohort and two validation cohorts were included in this study. Our prognostic DRGS can stratify CRC patients into two groups with different survival outcomes. A multivariate analysis suggested that the DRGS remained an independent prognostic factor and was significantly associated with poor prognosis in CRC. Furthermore, the C-index results of the DRGS showed its clinical superiority to Oncotype DX. Thus, it offers a significantly promising prognostic biomarker potential compared to the clinicopathological risk factors that are currently in use. The GSEA revealed that the metastasis-related pathways (i.e., angiogenesis, KRAS signaling, epithelial mesenchymal

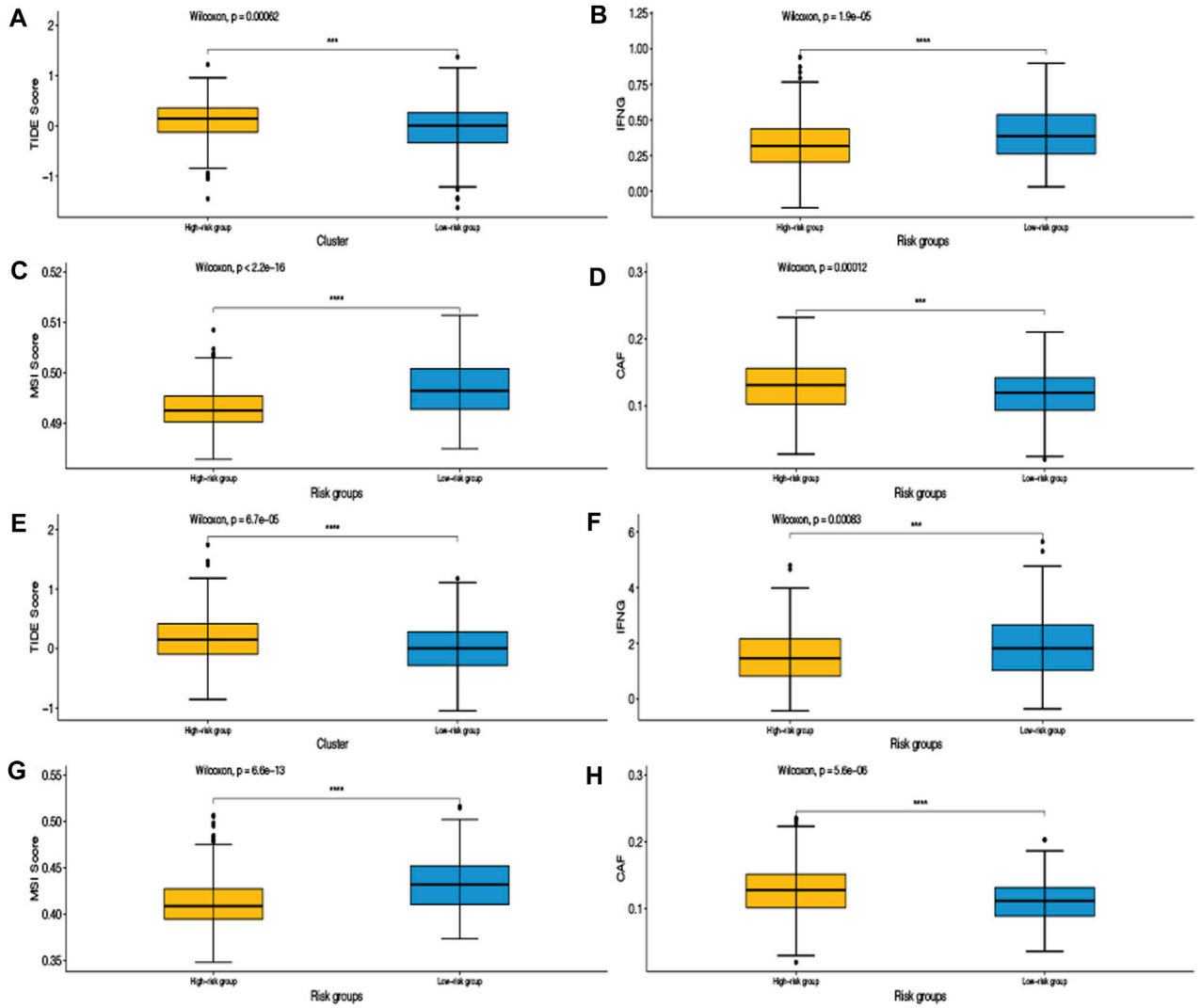


FIGURE 7 | Tumor immune dysfunction and exclusion (TIDE) algorithm was validated in the training set GSE39582 **(A,B,C,D)** and the validation set TCGA **(E,F,G,H)**.

transit, and myogenesis pathways) were enriched in the high-risk group, all of which were well known to play a crucial role in the progression and proliferation of CRC in numerous studies (Cooks et al., 2013; De Simone et al., 2015; Lu et al., 2016). Further studies are required to clarify the effects of DNA repair in order to identify more targets and improve the prognosis of CRC patients.

In order to further investigate the clinical application of our model, we divided the CRC cell lines in the GDSC database into the high-risk group and low-risk group according to the DRGS and analyzed the differences in chemotherapy response between the two groups. We can see that among the chemotherapeutic drugs such as oxaliplatin, 5-fluorouracil, and irinotecan, the IC50 of the cell line in the low-risk group was lower. It showed that the cell lines in the low-risk group were more sensitive to these three drugs. On the contrary, the cell lines in the high-risk group were more insensitive to these three chemotherapeutic drugs. This indicated that our model could predict the response of cell lines to chemotherapeutic agents. This may provide some guidance for clinical medication.

We knew that MSI/dMMR was widely considered as a potential biomarker for predicting ICB (Zhao et al., 2019). We wanted to know whether our model can predict immunotherapy, so we verified our model by using the data provided in the article that "Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma (ccRCC)" published in *Nature Medicine* (Braun et al., 2020). From the OS curve of the high- and low-risk groups, we could see that the OS of the high-risk group was worse in the ccRCC patients, and it suggested that our model can also well predict the OS of patients with ccRCC. When it comes to the ORR of immunotherapy, we can clearly see the ORR with CR + PR that had better OS in both the high-risk and low-risk groups. Notably, the OS of the low-risk group with CR + PR was the best.

This indicated that our model can potentially predict the response of patients to immunotherapy. Our model can be used to further identify cancer patients who are more suitable for immunotherapy.

To further demonstrate that our model can predict the response to immunotherapy, we used the TIDE algorithm for validation in the training and validation data sets. From the results, we can see that the TIDE score of the high-risk group was higher than that of the low-risk group, indicating that the high-risk group was less sensitive to immunotherapy than the low-risk group. That is to say, the low-risk group was more effective for immunotherapy. IFNG, produced by T cells in the immune system and natural killer cells, is a potent viral inhibitor (Jorgovanovic et al., 2020). MSI, caused by defects in MMR genes, is an important molecular marker for prognosis and the development of adjuvant treatment regimens in colorectal and other solid tumors (Boland and Goel, 2010). CAFs are a group of activated fibroblasts with significant heterogeneity and plasticity in the tumor microenvironment, which have significant tumor-promoting functions (Chen and Song, 2019). The low-risk group had higher IFNG, higher MSI score, and lower cancer CAFs amount, which showed that the immune landscape of the low-risk group was more active. The consistent results of the training and validation data sets not only proved the reliability and robustness of our model but also proved that our model can predict the response to immunotherapy, which may bring some clinical benefits to CRC patients.

As for how to apply the model to the clinic, we can detect these 11 genes for patients. Because it is a small panel of genes, it can avoid the waste of large medical resources and reduce the problem of high diagnostic cost for patients as much as possible. By detecting the 11 small panel genes, we calculated the risk score of patients and grouped them. With the help of the prediction model, not only patients can make more favorable choices for themselves but also doctors can make better clinical decisions according to the patient's risk score.

There are some limitations to our study. First, this is a retrospective study, although we validated the signature in independent data sets. In addition, the samples from primary tumor or metastatic disease may have inconsistent genetic heterogeneity, which could lead to sampling bias (NEJM Group, 2012; Mimori et al., 2018). In addition, systematic errors result from analyzing samples of disparate databases or the influence of measuring instruments, and not all batch effects can be eliminated based on their complexity. In verifying whether the model could predict immunotherapy response to CRC, we used immunotherapy data sets from ccRCC as there is currently a lack of data sets for public immunotherapy response to CRC. However, we also used the TIDE algorithm to further verify that our model can predict the immunotherapy benefit of patients. Therefore, we have sufficient evidence to prove that our model can predict the benefit of immunotherapy for patients. Although we investigated as many genes as possible, further clinical and pharmacological tests are needed to validate our results.

CONCLUSION

In summary, our work provides an accurate prognostic approach for estimating the survival outcomes of CRC patients. Further prospective studies are needed to evaluate the clinical application of this signature for the prognosis of CRC.

DATA AVAILABILITY STATEMENT

The data sets generated and analyzed during the current study are available in the TCGA cohort data downloaded from Broad GDAC Firehose (http://gdac.broadinstitute.org/). Other microarray data sets were acquired directly from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query).

AUTHOR CONTRIBUTIONS

M-YL and WW contributed equally to this study. M-YL, WW, XH and FG contributed to the study concept and design, acquisition, analysis, and interpretation of data and in drafting of the manuscript. M-EZ, DC, DF, C-HL, W-BK, Z-PH, XD, CH and QZ contributed to the data collections and manuscript reviews. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 82002221, FG), the Sun Yat-sen University 100 Top Talent Scholars Program – China (No. P20190217202203617, FG), Guangzhou Basic and Applied Basic Research Fund (No. 202102020820, FG), The Sixth Affiliated Hospital of Sun Yat sen University Startup Fund for Returnees (No. R20210217202501975, FG), Project funded by China Postdoctoral Science Foundation (No. 2020M683121, MZ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.872238/full#supplementary-material.

Supplementary Figure 1 | (A) Distribution of the DRGS risk score and its correlation to the survival status. **(B)** Time-dependent ROC analysis of overall survival for CRC patients at the time points of 5 and 10 y. **(C)** Kaplan–Meier curves comparing overall survival of patients within the low and high-risk groups. *P*-values were calculated using log-rank tests.

Supplementary Figure 2 | (A) GSEA showed several metastasis-related processes enriched in the high-risk group in the TCGA validation dataset. **(B)** GSEA showed several metastasis-related processes enriched in the high-risk group in the validation-2 cohort.

REFERENCES

Boland, C. R., and Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138 (6), 2073–2087. doi:10.1053/j.gastro.2009.12.064

Bouwman, P., and Jonkers, J. (2012). The Effects of Deregulated DNA Damage Signalling on Cancer Chemotherapy Response and Resistance. *Nat. Rev. Cancer* 12 (9), 587–598. doi:10.1038/nrc3342

Braun, D. A., Hou, Y., Bakouny, Z., Ficial, M., Sant' Angelo, M., Forman, J., et al. (2020). Interplay of Somatic Alterations and Immune Infiltration Modulates Response to PD-1 Blockade in Advanced clear Cell Renal Cell Carcinoma. *Nat. Med.* 26 (6), 909–918. doi:10.1038/s41591-020-0839-y

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Burdak-Rothkamm, S., and Rothkamm, K. (2021). DNA Damage Repair Deficiency and Synthetic Lethality for Cancer Treatment. *Trends Mol. Med.* 27, 91–92. doi:10.1016/j.molmed.2020.09.011

Chen, X., and Song, E. (2019). Turning Foes to Friends: Targeting Cancer-Associated Fibroblasts. *Nat. Rev. Drug Discov.* 18 (2), 99–115. doi:10.1038/s41573-018-0004-1

Cooks, T., Pateras, I. S., Tarcic, O., Solomon, H., Schetter, A. J., Wilder, S., et al. (2013). Mutant P53 Prolongs NF-Kb Activation and Promotes Chronic Inflammation and Inflammation-Associated Colorectal Cancer. *Cancer Cell* 23 (5), 634–646. doi:10.1016/j.ccr.2013.03.022

Das, V., Kalita, J., and Pal, M. (2017). Predictive and Prognostic Biomarkers in Colorectal Cancer: A Systematic Review of Recent Advances and Challenges. *Biomed. Pharmacother.* 87, 8–19. doi:10.1016/j.biopha.2016.12.064

De Rosa, M., Rega, D., Costabile, V., Duraturo, F., Niglio, A., Izzo, P., et al. (2016). The Biological Complexity of Colorectal Cancer: Insights into Biomarkers for Early Detection and Personalized Care. *Therap Adv. Gastroenterol.* 9 (6), 861–886. doi:10.1177/1756283X16659790

De Simone, V., Franzè, E., Ronchetti, G., Colantoni, A., Fantini, M. C., Di Fusco, D., et al. (2015). Th17-type Cytokines, IL-6 and TNF-α Synergistically Activate STAT3 and NF-kB to Promote Colorectal Cancer Cell Growth. *Oncogene* 34 (27), 3493–3503. doi:10.1038/onc.2014.286

Diboun, I., Wernisch, L., Orengo, C. A., and Koltzenburg, M. (2006). Microarray Analysis after RNA Amplification Can Detect Pronounced Differences in Gene Expression Using Limma. *BMC Genomics* 7, 252. doi:10.1186/1471-2164-7-252

Friedberg, E. C. (2001). How Nucleotide Excision Repair Protects against Cancer. *Nat. Rev. Cancer* 1 (1), 22–33. doi:10.1038/35094000

Fung, K. Y., Nice, E., Priebe, I., Belobrajdic, D., Phatak, A., Purins, L., et al. (2014). Colorectal Cancer Biomarkers: to Be or Not to Be? Cautionary Tales from a Road Well Travelled. *World J. Gastroenterol.* 20 (4), 888–898. doi:10.3748/wjg.v20.i4.888

Gao, F., Wang, W., Tan, M., Zhu, L., Zhang, Y., Fessler, E., et al. (2019). DeepCC: a Novel Deep Learning-Based Framework for Cancer Molecular Subtype Classification. *Oncogenesis* 8 (9), 44. doi:10.1038/s41389-019-0157-8

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (2015). The Consensus Molecular Subtypes of Colorectal Cancer. *Nat. Med.* 21 (11), 1350–1356. doi:10.1038/nm.3967

Iyama, T., and Wilson, D. M., 3rd (2013). DNA Repair Mechanisms in Dividing and Non-dividing Cells. *DNA Repair* 12 (8), 620–636. doi:10.1016/j.dnarep.2013.04.015

Jorgovanovic, D., Song, M., Wang, L., and Zhang, Y. (2020). Roles of IFN-γ in Tumor Progression and Regression: a Review. *Biomark Res.* 8, 49. doi:10.1186/s40364-020-00228-x

Kandimalla, R., Gao, F., Matsuyama, T., Ishikawa, T., Uetake, H., Takahashi, N., et al. (2018). Genome-wide Discovery and Identification of a Novel miRNA Signature for Recurrence Prediction in Stage II and III Colorectal Cancer. *Clin. Cancer Res.* 24 (16), 3867–3877. doi:10.1158/1078-0432.CCR-17-3236

Kandimalla, R., Ozawa, T., Gao, F., Wang, X., Goel, A., Nozawa, H., et al. (2019). Gene Expression Signature in Surgical Tissues and Endoscopic Biopsies Identifies High-Risk T1 Colorectal Cancers. *Gastroenterology* 156 (8), 2338–2341. e3. doi:10.1053/j.gastro.2019.02.027

Knijnenburg, T. A., Wang, L., Zimmermann, M. T., Chambwe, N., Gao, G. F., Cherniack, A. D., et al. (2018). Genomic and Molecular Landscape of DNA Damage Repair Deficiency across the Cancer Genome Atlas. *Cell Rep* 23 (1), 239–e6. doi:10.1016/j.celrep.2018.03.076

Lech, G., Slotwinski, R., Slodkowski, M., and Krasnodebski, I. W. (2016). Colorectal Cancer Tumour Markers and Biomarkers: Recent Therapeutic Advances. *World J. Gastroenterol.* 22 (5), 1745–1755. doi:10.3748/wjg.v22.i5.1745

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Lindahl, T., and Wood, R. D. (1999). Quality Control by DNA Repair. *Science* 286 (5446), 1897–1905. doi:10.1126/science.286.5446.1897

Lu, Y.-X., Ju, H.-Q., Wang, F., Chen, L.-Z., Wu, Q.-N., Sheng, H., et al. (2016). Inhibition of the NF-Kb Pathway by Nafamostat Mesilate Suppresses Colorectal Cancer Growth and Metastasis. *Cancer Lett.* 380 (1), 87–97. doi:10.1016/j.canlet.2016.06.014

Marcuello, M., Vymetalkova, V., Neves, R. P. L., Duran-Sanchon, S., Vedeld, H. M., Tham, E., et al. (2019). Circulating Biomarkers for Early Detection and Clinical Management of Colorectal Cancer. *Mol. Aspects Med.* 69, 107–122. doi:10.1016/j.mam.2019.06.002

Migliore, L., Migheli, F., Spisni, R., and Coppedè, F. (2011). Genetics, Cytogenetics, and Epigenetics of Colorectal Cancer. *J. Biomed. Biotechnol.* 2011, 1–19. doi:10.1155/2011/792362

Mimori, K., Saito, T., Niida, A., and Miyano, S. (2018). Cancer Evolution and Heterogeneity. *Ann. Gastroenterol. Surg.* 2 (5), 332–338. doi:10.1002/ags3.12182

Nejm Group (2012). Intratumor Heterogeneity and Branched Evolution. *New Engl. J. Med.* 366, 2132–2133. doi:10.1056/nejmc1204069

Ozawa, T., Kandimalla, R., Gao, F., Nozawa, H., Hata, K., Nagata, H., et al. (2018). A MicroRNA Signature Associated with Metastasis of T1 Colorectal Cancers to Lymph Nodes. *Gastroenterology* 154 (4), 844–848. doi:10.1053/j.gastro.2017.11.275

Phipps, A. I., Limburg, P. J., Baron, J. A., Burnett-Hartman, A. N., Weisenberger, D. J., Laird, P. W., et al. (2015). Association between Molecular Subtypes of Colorectal Cancer and Patient Survival. *Gastroenterology* 148 (1), 77–87. doi:10.1053/j.gastro.2014.09.038

Schröder, M. S., Culhane, A. C., Quackenbush, J., and Haibe-Kains, B. (2011). Survcomp: an R/Bioconductor Package for Performance Assessment and Comparison of Survival Models. *Bioinformatics* 27 (22), 3206–3208. doi:10.1093/bioinformatics/btr511

Shah, R., Jones, E., Vidart, V., Kuppen, P. J. K., Conti, J. A., and Francis, N. K. (2014). Biomarkers for Early Detection of Colorectal Cancer and Polyps: Systematic Review. *Cancer Epidemiol. Biomarkers Prev.* 23, 1712–1728. doi:10.1158/1055-9965.EPI-14-0412

Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer Statistics, 2016. *CA: A Cancer J. Clinicians* 66 (1), 7–30. doi:10.3322/caac.21332

Siegel, R. L., Fedewa, S. A., Anderson, W. F., Miller, K. D., Ma, J., Rosenberg, P. S., et al. (2017). Colorectal Cancer Incidence Patterns in the United States, 1974-2013. *J. Natl. Cancer Inst.* 109 (8), djw322. doi:10.1093/jnci/djw322

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model. *Stat. Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

Wang, X., Terfve, C., Rose, J. C., and Markowetz, F. (2011). HTSanalyzeR: an R/Bioconductor Package for Integrated Network Analysis of High-Throughput Screens. *Bioinformatics* 27 (6), 879–880. doi:10.1093/bioinformatics/btr028

Zhao, P., Li, L., Jiang, X., and Li, Q. (2019). Mismatch Repair Deficiency/microsatellite Instability-High as a Predictor for Anti-PD-1/pd-L1 Immunotherapy Efficacy. *J. Hematol. Oncol.* 12 (1), 54. doi:10.1186/s13045-019-0738-1

Zheng, Z.-X., Zheng, R.-S., Zhang, S.-W., and Chen, W.-Q. (2014). Colorectal Cancer Incidence and Mortality in China, 2010. *Asian Pac. J. Cancer Prev.* 15 (19), 8455–8460. doi:10.7314/apjcp.2014.15.19.8455

Zinovkina, L. A. (2018). Mechanisms of Mitochondrial DNA Repair in Mammals. *Biochem. Mosc.* 83 (3), 233–249. doi:10.1134/s0006297918030045

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 Lv, Wang, Zhong, Cai, Fan, Li, Kou, Huang, Duan, Hu, Zhu, He and Gao. 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 Physiology

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fphys.2022.880981

Endoscopy-Based Deep Convolutional Neural Network Predicts Response to Neoadjuvant Treatment for Locally Advanced Rectal Cancer

Xijie Chen[1,2‡], Junguo Chen[1,3‡], Xiaosheng He[1,3‡], Liang Xu[1,4], Wei Liu[1,6], Dezheng Lin[1,6], Yuxuan Luo[5], Yue Feng[5], Lei Lian[1,2], Jiancong Hu[1,6,7]* and Ping Lan[1,3]**

[1]Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2]Department of Gastrointestinal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [3]Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [4]Department of Pathology, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [5]Tianjin Economic-Technological Development Area, Yujin Digestive Health Industry Research Institute, Tianjin, China, [6]Department of Endoscopic Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [7]Department of Network Management, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

OPEN ACCESS

Edited by:
Wei Wang,
Huzhou Maternity and Child Health Care Hospital, China

Reviewed by:
Longhui Qiu,
University of California, San Francisco, United States
Jianwei Hu,
Fudan University, China

***Correspondence:**
Lei Lian
lianlei2@mail.sysu.edu.cn
Jiancong Hu
hujianc@mail.sysu.edu.cn
Ping Lan
lanping@mail.sysu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

Received: 22 February 2022
Accepted: 11 March 2022
Published: 27 April 2022

Citation:
Chen X, Chen J, He X, Xu L, Liu W, Lin D, Luo Y, Feng Y, Lian L, Hu J and Lan P (2022) Endoscopy-Based Deep Convolutional Neural Network Predicts Response to Neoadjuvant Treatment for Locally Advanced Rectal Cancer. Front. Physiol. 13:880981. doi: 10.3389/fphys.2022.880981

Background and Aims: Although the wait and watch (W&W) strategy is a treatment choice for locally advanced rectal cancer (LARC) patients who achieve clinical complete response (cCR) after neoadjuvant therapy (NT), the issue on consistency between cCR and pathological CR (pCR) remains unsettled. Herein, we aimed to develop a deep convolutional neural network (DCNN) model using endoscopic images of LARC patients after NT to distinguish tumor regression grade (TRG) 0 from non-TRG0, thus providing strength in identifying surgery candidates.

Methods: A total of 1000 LARC patients (6,939 endoscopic images) who underwent radical surgery after NT from April 2013 to April 2021 at the Sixth Affiliated Hospital, Sun Yat-sen University were retrospectively included in our study. Patients were divided into three cohorts in chronological order: the training set for constructing the model, the validation set, and the independent test set for validating its predictive capability. Besides, we compared the model's performance with that of three endoscopists on a class-balanced, randomly selected subset of 20 patients' LARC images (10 TRG0 patients with 70 images and 10 non-TRG0 patients with 72 images). The measures used to evaluate the efficacy for identifying TRG0 included overall accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUROC).

Results: There were 219 (21.9%) cases of TRG0 in the included patients. The constructed DCNN model in the training set obtained an excellent performance with good accuracy of 94.21%, specificity of 94.39%, NPV of 98.11%, and AUROC of 0.94. The validation set showed accuracy, specificity, NPV, and AUROC of 92.13%, 93.04%, 96.69%, and 0.95, respectively; the corresponding values in the independent set were 87.14%, 92.98%,

91.37%, and 0.77, respectively. In the reader study, the model outperformed the three experienced endoscopists with an AUROC of 0.85.

Conclusions: The proposed DCNN model achieved high specificity and NPV in detecting TRG0 LARC tumors after NT, with a better performance than experienced endoscopists. As a supplement to radiological images, this model may serve as a useful tool for identifying surgery candidates in LARC patients after NT.

Keywords: treatment response, endoscopy, deep convolutional neural network, rectal cancer, neoadjuvant therapy

INTRODUCTION

The incorporation of neoadjuvant therapy (NT) can help downstage and downsize primary rectal cancer, prevent local recurrence, and increase the possibility of sphincter preservation (Maas et al., 2010; Al-Sukhni et al., 2016). In contrast, radical surgery is invasive and can lead to severe complications, including permanent stoma, sexual, bladder, and bowel dysfunction. The perioperative mortality rates of radical surgical resection of LARC are as high as 2–5% (Borowski et al., 2010; Marijnen, 2015). Hence, LARC patients who have a clinical complete response (cCR, highly suspected as pathological CR [pCR]) after NT sometimes opt for a wait and watch (W&W) strategy. Indeed, some studies have found that among LARC patients who had a cCR after NT, those who opted for a W&W strategy had the same survival rate as those who underwent radical surgery (Coraglio et al., 2020; Simpson et al., 2020).

However, there is no standardized definition of cCR (van der Valk et al., 2018). High rates of residual tumor cells have been found in tumor specimens resected from cCR patients, which indicates poor consistency between cCR and pCR (Hiotis et al., 2002; Guillem et al., 2005). Among patients with cCR but not pCR who opt for the W&W strategy, a burst local recurrence within 2 years followed by difficult salvage surgery with more complex complications are foreseeable problems. Thus, current evidence suggests that the W&W strategy should be applied with caution.

Deep convolutional neural network (DCNN), a branch of artificial intelligence (AI), has a unique capacity for the integration of high-dimensional data, and is well suited to the medical field, with promising applications in capturing the features of deep layers (Li et al., 2018; Xie et al., 2020; Jiang et al., 2021). The DCNN network mimics the structure and activity of the brain neurons, which is logically in line with human thinking, and optimized on this basis. A DCNN model enables machines to train various given images derived from different inspection equipment and extracts specific clinical characteristics using a backpropagation algorithm. Based on these clinical characteristics, the machine is able to make diagnosis from newly acquired clinical images prospectively. Few studies have investigated the utility of DCNNs for the analysis of endoscopic images to detect TRG. The present study mainly aimed to develop a DCNN model to evaluate TRG0 using endoscopic images in LARC patients after NT, and compare the model's performance with that of experienced gastrointestinal endoscopists on the same test set. In this way, we hoped to avoid the incorrect application of the W&W strategy, which would result in patients missing their optimal time window for surgery.

MATERIALS AND METHODS

Ethics Statement

This study was conducted in concordance with the ethical standards of the World Medical Association, the tenets of the Declaration of Helsinki, and the Ethical Guidelines for Clinical Research. In addition, the study was approved by the institutional review board of the Sixth Affiliated Hospital, Sun Yat-sen University (no. 2021ZSLYEC-063). Informed consent was not required because pre-existing data were used.

Patient Selection

We initially evaluated 1,103 consecutive patients who were pathologically diagnosed with rectal cancer. All the patients underwent radical surgery after NT between April 2013 and April 2021 in the Sixth Affiliated Hospital, Sun Yat-sen University.

Patients who received neoadjuvant chemoradiotherapy were given long-course radiotherapy of 50 Gy in 25 fractions or short-course radiotherapy of 5 Gy once a day for 5 days to the clinical target volume. The concurrent chemotherapy regimen was mainly based on oral/intravenous 5-fluorouracil, or combined with oxaliplatin/irinotecan. Patients treated with neoadjuvant chemotherapy alone shared the same chemotherapy regimen with the former.

The inclusion criteria were as follows: 1) patients, primary T3-4/N+ rectal cancer with post-treatment restaging endoscopy data stored in our center, 2) intervention, NT followed by radical surgery, and 3) outcomes, tumor regression grade (TRG) assessments (derived from surgical pathology reports) available. We excluded patients who had 1) ambiguous endoscopy images (e.g., low-resolution, under-focus, etc.), 2) bleeding on the tumor surface by endoscopic procedure or inadequate bowel preparation (Boston Bowel Preparation Score lower than 2 for the examined regions of the rectum), which blurs the tumor surface and hinders feature extraction, 3) endoscopic imaging with operating forceps, 4) insufficient NT courses (withdrawal from planned chemoradiotherapy protocol), 5) simultaneous colorectal cancer, 6) familial adenomatous polyps, Lynch syndrome, ulcerative colitis, or other diseases with genetic susceptibility to colorectal cancer, or 7) missing data. The raw screening process of the images was

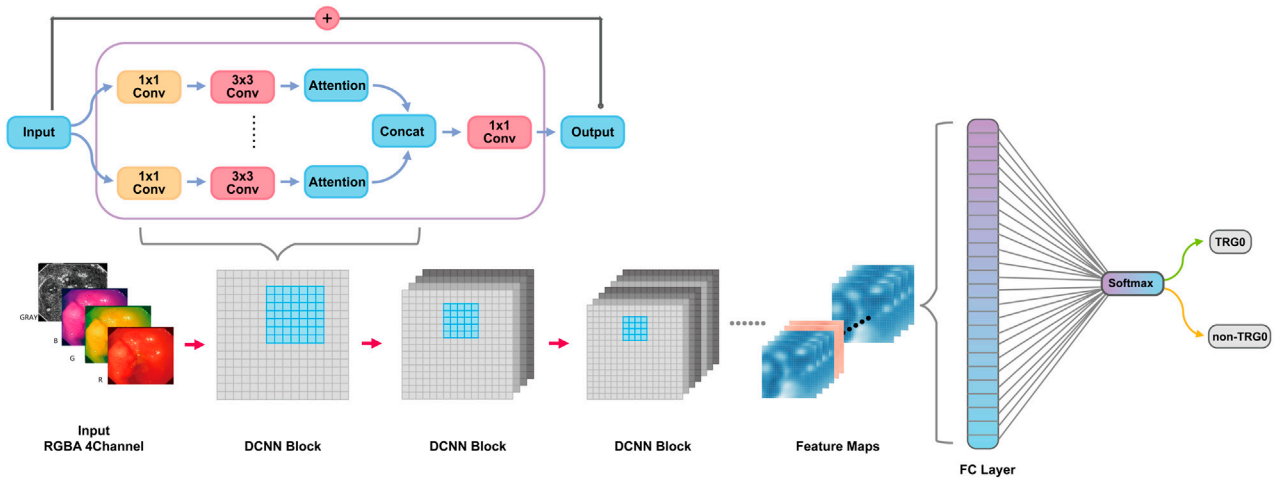


FIGURE 1 | The key architecture diagram of the DCNN model. DCNN, deep convolutional neural network.

evaluated by two experienced endoscopists; when disagreement occurred, the opinion of the third senior investigator was requested. To create a more homogeneous model, at least four images which met the prespecified requirements for image quality of each person were required.

Data Preparation

Preoperative, unamplified, white-light endoscopic images with diagnostic reports and postsurgical pathology reports were collected from the digital image acquisition and pathology report systems, respectively. Restaging endoscopic images were collected 6–8 weeks after finishing NT. The interval time from finishing NT to surgery was described as 6–12 weeks. We extracted tumor information such as differentiation and TRG. As described by the AJCC 8th edition staging system (Weiser, 2018), TRGs were defined as follows: TRG0 (complete response), no viable cancer cells; TRG1 (moderate response), single or small groups of cancer cells; TRG2 (minimal response), residual cancer outgrown by fibrosis; and TRG3 (poor response), minimal or no tumor kill and extensive residual cancer. Patients' demographic information, including sex, age, and tumor biomarkers, was acquired from electronic or paper health records.

Outcome and Group Assignment

The primary outcome of this study was to develop a DCNN model to discriminate TRG0 from non-TRG0 (TRG1—TRG3) among LARC patients who had received NT. We randomly divided patients treated between April 2013 and October 2020 into a training set and a validation set at a ratio of 9:1; the training set was used for constructing the DCNN model, while the validation set was used to examine its predicting capability. Patients treated between November 2020 and April 2021 were assigned to a completely independent test set to externally evaluate the model's performance in tumor response prediction. Receiver operating characteristic (ROC) curves were plotted to detect the discriminative power, which was quantified using the area under the ROC curve (AUROC).

To compare the performance of DCNN model with that of veteran endoscopists, we did a manual review study (reader study) in which three endoscopists reviewed the same test set of 20 patients' LARC images after NT (10 TRG0 patients with 70 images and 10 non TRG-0 patients with 72 images), which were class-balanced and randomly selected from the independent test set. Each endoscopist had more than 3 years of work experience and performed over 5,000 endoscopy examinations.

DCNN Model Construction

After image preprocessing (details shown in the supplementary method section, **Supplementary Figure S1**), we tailored a modification of the ResNeSt-50 variant, an existing high-performance neural network, and preprocessed the pathological data to obtain more effective features. DCNN is famous for its powerful ability of feature extraction and classification and recognition. In this study, DCNN network named ResNeSt-50 was used to identify endoscopic images as TRG0 or non-TRG0. We overlayed the RGB three channels of the image with the gray image with enhanced edge features, and synthesize four channels as input information. DCNN used each internal block to complete feature extraction, and continuously upgraded the dimension in this process to extract some abstract high-level features, and finally form several feature maps. At the end of DCNN, these feature maps would be transformed into a full connection layer (FC layer), and the regression operation would be carried out by using softmax function to obtain the results of TRG0 or non-TRG0 (**Figure 1**). To visualize the characteristics of the lesions, we created a heat map (**Figure 2**) to better display the features, making the network more sensitive to the local details of the space and image. We used label smoothing technology to prevent overfitting, thereby increasing the fault tolerance rate, generating a better calibration network, and ultimately incubating a more accurate

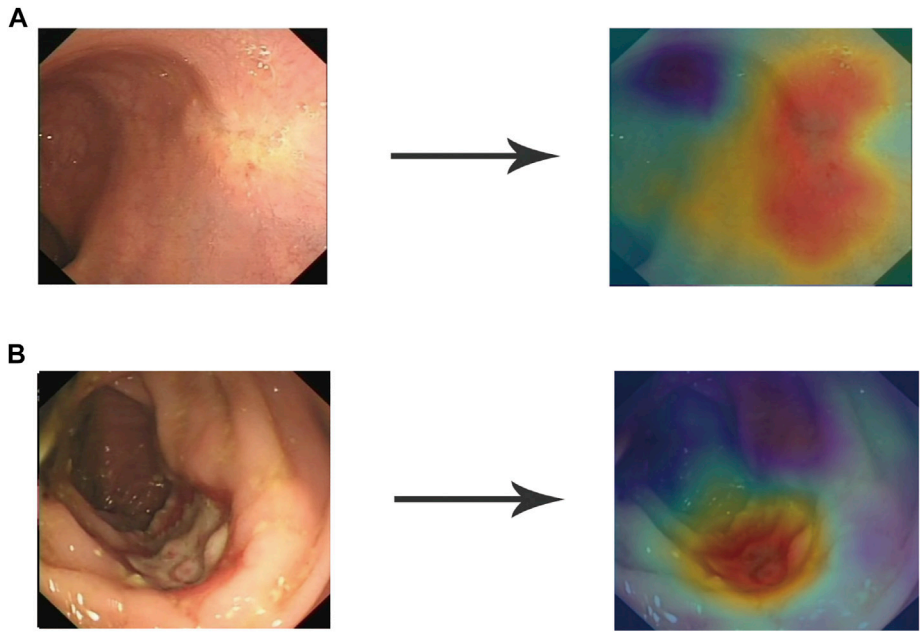


FIGURE 2 | Heat map to visualize the DCNN model. The heat map is mainly composed of red tones and blue tones. The red tones reveal the very region of the input image that activates the category (TRG0 or non-TRG0), which is what we are interested in, while the blue tones are the regions of non-interest. The darker the red tone, the more important the region. **(A)** Representative image of TRG0. **(B)** Representative image of non-TRG0. DCNN, deep convolutional neural network.

prediction model on invisible data. In addition, a weighting method was used to eliminate the bias caused by imbalanced data distribution; the formula was as follows: category weight = (total sample–current category samples)/total sample. Finally, after pre-training by the ImageNet, we used our data for fine-tuned learning and model training. The initial learning rate was 1e–4, and the momentum was 0.9; we stopped the training at 300 epochs.

Evaluation Indicators and Statistical Analysis

Evaluation indicators, including AUROC, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The 95% confidence interval (CI) was calculated using a two-sided exact binomial test using the Clopper-Pearson interval. For clinicopathological data, continuous random variables with normal distributions were presented as mean and standard deviation, and non-normally distributed variables were presented as median and interquartile range. Categorical variables were presented as frequencies and percentages. For the sake of actual clinical practice, continuous random variables were transformed into categorical variables when appropriate. The chi-square test was used for qualitative data. A stepwise binary logistic regression was performed to identify factors that were independently associated with our outcome of interest. The Pytorch (version 1.7.1) deep learning platform was employed for training and validating our DCNN model. For the reader study, TRG0 prediction agreement among the endoscopists was calculated using Fleiss' kappa. All the analyses were carried out using Python (version 3.7, Python Software Foundation, Wilmington, DE) and IBM SPSS statistics (version 26.0, IBM Corp., New York, USA). A two-sided p value of less than 0.05 was deemed statistically significant.

RESULTS

Patient Selection and Demographic Characteristics

A total of 1,000 LARC patients with 6,939 images obtained between April 2013 and April 2021 were included in this study (**Figure 3**). Among them, 219 patients had TRG0, accounting for 21.9% of the total population. The training and validation sets included 930 patients (6,500 endoscopic images), of whom 206 patients (1,433 images) were TRG0, and 724 patients (5,067 images) were non- TRG0. The independent set comprised 70 patients (439 images), among which 13 (18.6%) patients with 83 images were TRG0, and 57 patients (356 images) were non- TRG0. The characteristics of the two cohorts were comparable. The median age of the patients was 57 years, and male patients accounted for majority of subjects (~ 70%). All patients received neoadjuvant chemotherapy, and >40% of patients were administered neoadjuvant radiotherapy. Most of the cancers were moderately or poorly differentiated, and located in the middle or lower rectum. Obvious tumor regression could be observed in terms of the T/N stage and carcinoembryonic antigen (CEA) level after NT (**Table 1**). And we found that patients who were less than 50 years old, had received

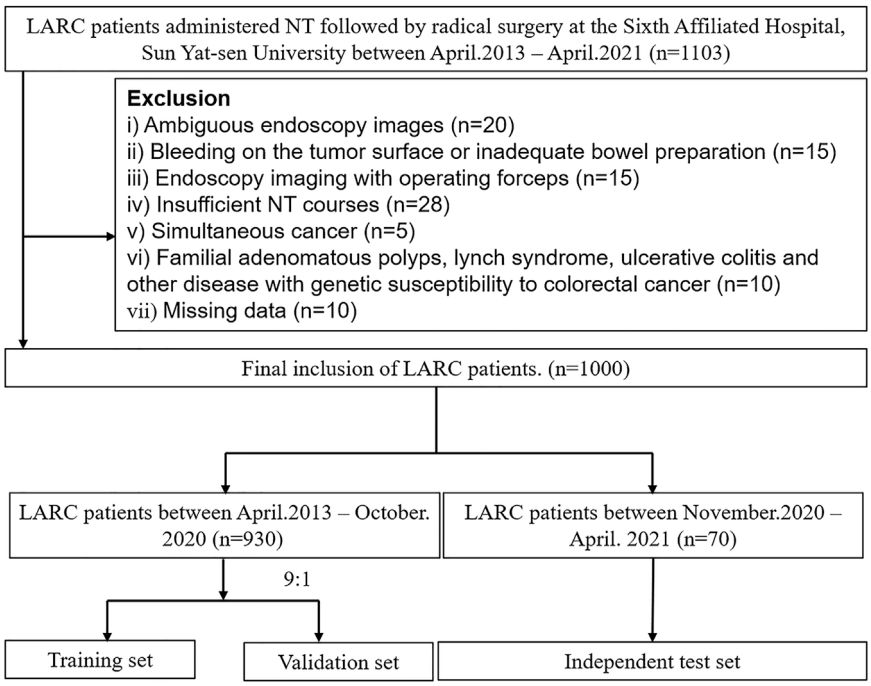


FIGURE 3 | Flow chart of the study. LARC, locally advanced rectal cancer; NT, neoadjuvant therapy.

neoadjuvant radiotherapy, had well-differentiated tumors, or had negative preoperative serum CEA levels (<5 μg/ml) were more likely to achieve TRG0. (Supplementary result section, **Supplementary Table S1**).

Modified DCNN Model Showed Good Performance in Predicting TRG Classification

Using the modified ResNeSt-50 variant, we constructed a DCNN model classifier to accurately predict tumor response in LARC patients who had received NT. In the training cohort, the model successfully recognized 4,308 images for non-TRG0 and 1,205 images for TRG0, with 83 images misdiagnosed as non-TRG0 and 256 images misdiagnosed as TRG0, yielding an AUROC of 0.94 (95% CI: 0.93, 0.95), suggesting no departure from the perfect fit (**Figure 4A**), and the overall accuracy of the model was 94.21% (95% CI: 0.94, 0.95), with a sensitivity, specificity, PPV, and NPV of 93.56% (95% CI: 0.92, 0.95), 94.39% (95% CI: 0.94, 0.95), 82.48% (95% CI: 0.80, 0.84), and 98.11% (95% CI: 0.98, 0.99), respectively. In the validation cohort, the model successfully recognized 468 images for non-TRG0 and 129 images for TRG0; with 16 images misdiagnosed as non-TRG0 and 35 images misdiagnosed as TRG0, yielding an AUROC of 0.95 (95% CI: 0.92, 0.98) (**Figure 4B**), and the overall accuracy was 92.13% (95% CI: 0.90, 0.94), with a sensitivity, specificity, PPV, and NPV of 88.97% (95% CI: 0.83, 0.93), 93.04% (95% CI: 0.90, 0.95), 78.66% (95% CI: 0.72, 0.84), and 96.70% (95% CI: 0.95, 0.98), respectively. In the independent test set, the model successfully recognized 53 patients for non-TRG0 and 8 patients for TRG0, with 4 patients misdiagnosed as non-TRG0 and 5 patients misdiagnosed as TRG0, yielding an AUROC of 0.77 (95% CI: 0.65, 0.93) (**Figure 4C**), and the overall accuracy was 87.14% (95% CI: 0.76, 0.94), with a sensitivity, specificity, PPV, and NPV of 61.53% (95% CI: 0.32, 0.85), 92.98% (95% CI: 0.82, 0.98), 66.67% (95% CI: 0.35, 0.89), and 91.37% (95% CI: 0.80, 0.97), respectively, indicating perfect reproducibility of the DCNN model (**Table 2**). We tested the performance of the DCNN model by using Chi-square and univariate logistic regression methods, and both showed that the model predicted actual events well (**Table 3**).

For the reader study, the model successfully recognized 9 patients for non-TRG0 and 8 patients for TRG0, with 2 patients misdiagnosed as non-TRG0 and 1 patient misdiagnosed as TRG0. Thus, the DCNN model achieved an accuracy of 85% (95% CI: 0.64-0.95), sensitivity of 80% (95% CI: 0.49-0.94), specificity of 90% (95% CI: 0.60-0.98), PPV of 88.89% (95% CI: 0.57-0.98), NPV of 81.82% (95% CI: 0.52-0.95), and AUROC of 0.85 (95% CI: 0.68-1). The Fleiss' kappa value was 0.722 ($p < 0.01$), which showed significant inter- and intra-observer variability among experts on the assessment of TRG status based on the colonoscopic images. However, the DCNN model performed significantly better than the three endoscopists almost in all evaluating indicators (**Table 4**), and the performance of all three endoscopists was below the model's ROC (**Figure 5**).

DISCUSSION

To the best of our knowledge, the present study is the first to develop an AI model by using endoscopic images from LARC patients after NT for the prediction of TRG0. Our DCNN model

TABLE 1 | Characteristics of locally advanced rectal cancer patients.

	Training Set + Validation Set	Independent Test Set	*p*
TRG	—	—	0.64
0	206 (22.2%)	13 (18.6%)	—
1	200 (21.5%)	12 (17.1%)	—
2	415 (44.6%)	36 (51.4%)	—
3	109 (11.7%)	9 (12.9%)	—
Age	57 (47–64)	57.5 (50–64)	0.73
Sex	—	—	0.20
male	669 (71.9%)	47 (67.1)	—
female	261 (28.1%)	12 (32.9)	—
BMI	22.6 (20.5–24.8)	22.6 (20.8–24.1)	0.76
Neoadjuvant chemotherapy	—	—	1
yes	930 (100%)	70 (100%)	—
no	0 (0)	0 (0)	—
Neoadjuvant radiotherapy	—	—	0.74
yes	380 (40.9%)	30 (42.9%)	—
no	550 (59.1%)	40 (57.1%)	—
Differentiation	—	—	0.36
well	259 (27.8%)	14 (20.0%)	—
moderate	592 (63.7%)	49 (70.0%)	—
poor	79 (8.5%)	7 (10.0%)	—
Pre-T[a]	—	—	0.01[b]
2	34 (3.7%)	0 (0)	—
3	573 (61.6%)	36 (51.4%)	—
4	180 (19.4%)	22 (31.4%)	—
Pre-N[a]	—	—	0.25
0	158 (17.0%)	17 (24.3%)	—
1	297 (31.9%)	18 (25.7%)	—
2	315 (34.0%)	23 (32.8%)	—
ypT	—	—	0.56
0	208 (22.4%)	13 (18.6%)	—
1	67 (7.2%)	5 (7.1%)	—
2	223 (24.0%)	17 (24.3%)	—
3	417 (44.8%)	35 (50.0%)	—
4	15 (1.6%)	0 (0%)	—
ypN	—	—	0.56
0	706 (75.9%)	57 (81.4%)	—
1	158 (17.0%)	9 (12.9%)	—
2	66 (7.1%)	4 (5.7%)	—
initial CEA	4.5 (2.4–10.5)	5.0 (2.5–12.0)	0.26
preoperative CEA	2.78 (1.90–4.71)	2.4 (1.7–3.2)	0.73
Distal margin from the anal verge/mm[b]	52 (36–72)	50 (32–65.8)	0.17

[a]*Incomplete data.*
[b]*Significant different.*
TRG, tumor response grade; BMI, body mass index; pre-T, Pretreatment T stage.
Pre-N, Pretreatment N stage; CEA, carcinoembryonic antigen.

achieved good accuracy, high sensitivity, specificity, and NPV, and was proved to outperform experienced endoscopists. This model has the potential to serve as a robust supplementary tool to radiographic examinations for precisely selecting surgery candidates for LARC patients after NT.

NT followed by radical surgery is the standard of care for LARC. However, the fibrosis in the rectum and surrounding mesorectal tissue induced by NT increases the difficulty of radical surgery. Moreover, radical surgery is associated with mortality and morbidity such as permanent stoma, and can be physically, mentally, and emotionally traumatic to patients (Hupkens et al., 2017). Among LARC patients who achieve cCR, the W&W strategy or surgery could be chosen through the "share-decision making" policy in case of notifying specific risks. Some studies indicated that the prognosis of those who request the W&W strategy has not been found to be inferior to the prognosis of those who undergo radical surgery, and the absence of surgery greatly improves their quality of life (Sauer et al., 2004; Roh et al., 2009; Sebag-Montefiore et al., 2009; Valentini et al., 2011; Ortholan et al., 2012; Valentini et al., 2015). However, no standard criteria for cCR are available, cCR rates vary greatly from 10 to 78%, while the ensuing local recurrence or distant metastasis rates range from 7 to 33% (van der Valk et al., 2018; Pang et al., 2019; Asoglu et al., 2020; Pinto et al., 2020). This is partially attributable to an inflated false-positive rate due to estimations based on different standards. Therefore, enrollment in the W&W strategy without establishing proper surveillance protocols and salvage management might

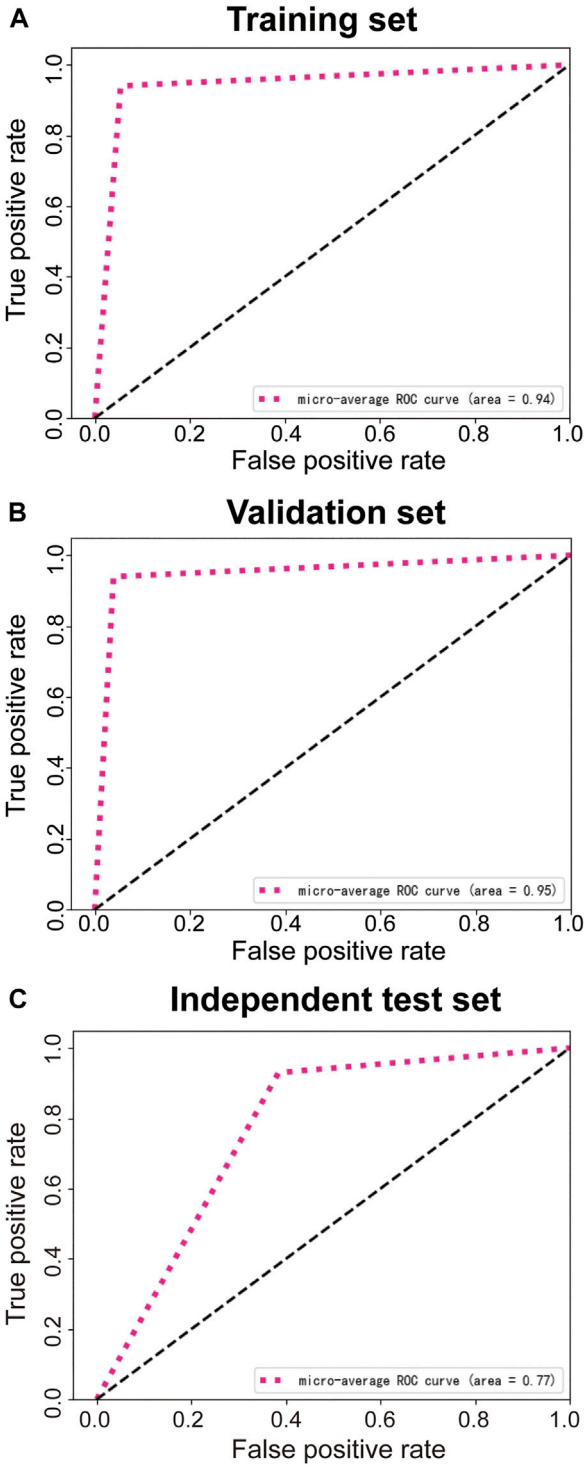


FIGURE 4 | ROC curves of the training set **(A)**, validation set **(B)**, and independent test set **(C)**. ROC, receiver operating characteristic.

result in dismal outcomes. The current study on the DCNN model aimed to accurately identify candidates for surgery and reduce the unsuitable application of watchful waiting.

Computerized tomography (CT), MRI, ultrasonography, digital rectal examination, biopsy examination, and assessment of certain morphological features on endoscopy have been recommended as modalities for tumor-response assessment during NT (Glynne-Jones et al., 2017; Dattani et al., 2018; van der Valk et al., 2018), but none of them can precisely differentiate between patients who require watchful waiting and those who

TABLE 2 | Efficacy of the DCCN model.

	Sensitivity	Specificity	PPV	NPV	Accuracy	AUROC
Training set	93.56% (95% CI: 0.92, 0.95)	94.39% (95% CI: 0.94, 0.95)	82.48% (95% CI: 0.80, 0.84)	98.11% (95% CI: 0.98, 0.99)	94.21% (95% CI: 0.94, 0.95)	0.94 (95% CI: 0.93, 0.95)
Validation set	88.97% (95% CI: 0.83, 0.93)	93.04% (95% CI: 0.90, 0.95)	78.66% (95% CI: 0.72, 0.84)	96.69% (95% CI: 0.95, 0.98)	92.13% (95% CI: 0.90, 0.94)	0.95 (95% CI: 0.92, 0.98)
Independent test set	61.53% (95% CI: 0.32, 0.85)	92.98% (95% CI: 0.82, 0.98)	66.67% (95% CI: 0.35, 0.89)	91.37% (95% CI: 0.80, 0.97)	87.14% (95% CI: 0.76, 0.94)	0.77 (95% CI: 0.65, 0.93)

CI, confidence interval.
DCNN, deep convolutional neural network.

TABLE 3 | Correlation of the DCNN model and actual events.

	Chi-Square		Univariate Logistic Regression		
	$\chi 2$	*P*	OR	95% CI	*P*
DCNN model (training set)	3876.33	<0.01[a]	167.07	132.08–211.32	<0.01[a]
Validation set	388.48	<0.01[a]	108.62	56.72–208.04	<0.01[a]
Independent test set	24.19	<0.01[a]	0.04	0.01–0.19	<0.01[a]

[a]*Significant different.*
DCNN, deep convolutional neural network.

TABLE 4 | Reader study.

	Sensitivity	Specificity	PPV	NPV	Accuracy	AUROC
DCNN model	80% (95% CI: 0.49, 0.94)	90% (95% CI: 0.60, 0.98)	88.89% (95% CI: 0.57, 0.98)	81.82% (95% CI: 0.52, 0.95)	85% (95% CI: 0.64, 0.95)	0.85 (95% CI: 0.69, 1)
Endoscopist 1	40% (95% CI: 0.17, 0.69)	70% (95% CI: 0.40, 0.89)	57% (95% CI: 0.25, 0.84)	53.8% (95% CI: 0.29, 0.77)	55% (95% CI: 0.34, 0.74)	0.55 (95% CI: 0.33, 0.77)
Endoscopist 3	80% (95% CI: 0.49, 0.94)	80% (95% CI: 0.49, 0.94)	80% (95% CI: 0.49, 0.94)	80% (95% CI: 0.49, 0.94)	80% (95% CI: 0.58, 0.92)	0.70 (95% CI: 0.49, 0.91)
Endoscopist 2	50% (95% CI: 0.24, 0.76)	80% (95% CI: 0.49, 0.94)	71.4% (95% CI: 0.36, 0.92)	61.5% (95% CI: 0.36, 0.82)	65% (95% CI: 0.43, 0.82)	0.65 (95% CI: 0.44, 0.86)

CI, confidence interval.

require radical surgery. Liu et al. (Liu et al., 2018) assessed 124 LARC patients, and found that only 25% sensitivity could be achieved regardless of the modality used to assess tumor response after NT, namely, MRI, ultrasonography, and endoscopic mucosal integrity. Although these techniques were associated with specificities of >90%, their use resulted in the recommendation of unnecessary surgery for at least 75% of pCR patients. Although some clinical characteristics were reported to be useful TRG0 predictors, there were significant inter-observer variability among three endoscopists in the reader study. Similarly, a study by van der Sande et al. (van der Sande et al., 2021) investigating the utility of endoscopic findings for TRG evaluation found that a flat scar was the feature most predictive of CR, with a PPV of 70–80%. However, due to the subjective nature of visual observation, only poor-to-moderate inter-observer agreement could be achieved among endoscopists. The potential of computer-aided systems to assist clinicians in diagnosing and evaluating gastrointestinal tumor lesions would help to deal with human subjectivity. Our findings suggest that the present deep learning model could provide added value as an automated screening tool of patient triage for confirmatory testing.

Recently, several radiomics and AI studies based on MRI have been conducted (Tang et al., 2019; Zhang et al., 2020). However, no radiomics model has yet been used in clinical practice. Factors that hamper the clinical application of radiomics are the lack of a unified standard for feature extraction and lack of evidence for the generalizability of the models across different MR scanners and different magnetic field strengths. In contrast, we constructed a DCNN model to predict TRG0 by using endoscopic images. The TRG0 rates in the training, validation, and independent test sets were 22.2, 20.4, and 18.6%, respectively, which are consistent with previous studies (Dattani et al., 2018; van der Valk et al., 2018). When developing this model, we focused on whether its sensitivity, specificity, and NPV were high enough to identify patients who indeed required surgery and to maximally recognize TRG0. As a result, although the diagnostic accuracy of the model may be affected due to various quality of optical imaging in different time periods as we divided the patients into three cohorts in chronological order, the model in the independent test set still accurately identified 8 CR and 53 non-CR patients; among the remaining patients, 5 were misdiagnosed with CR, and 4 were misdiagnosed with non-CR, yielding an accuracy of 87.14%, a specificity of 92.98%, and NPV of

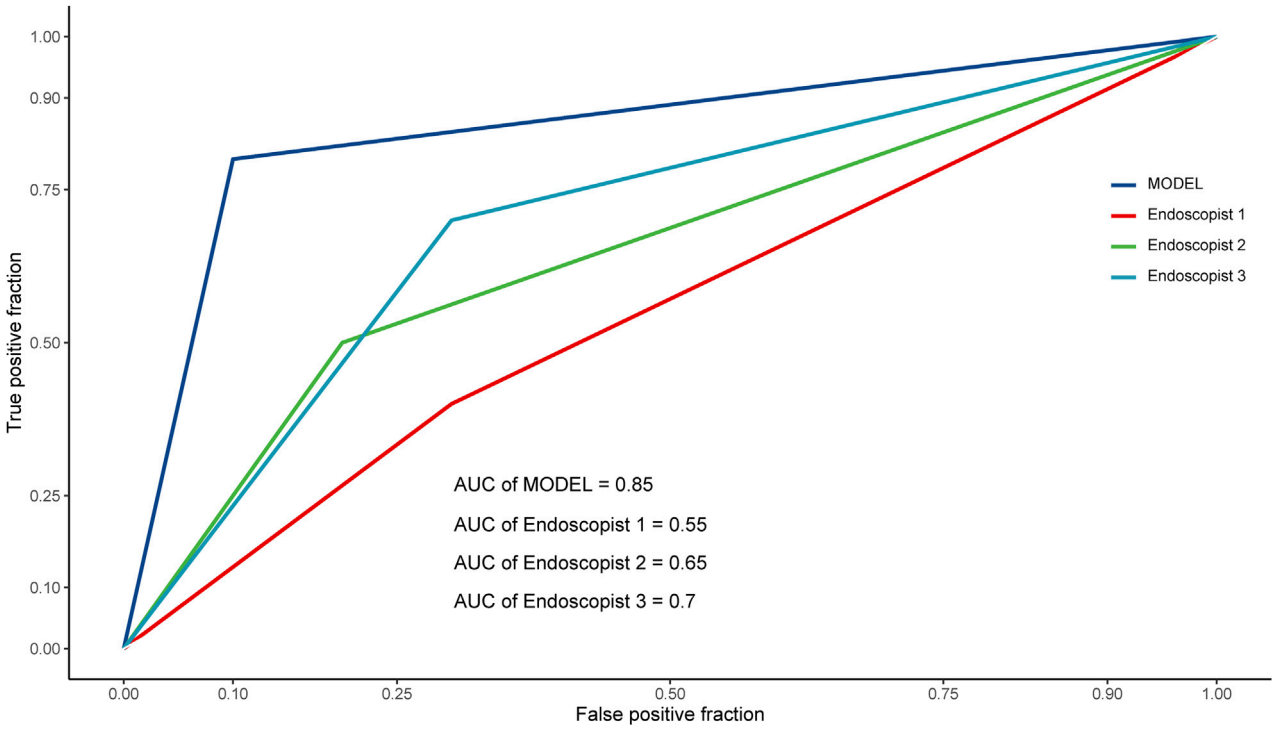


FIGURE 5 | ROC curves of the reader study. ROC, receiver operating characteristic; AUC (AUROC), area under the receiver operating characteristic curve.

91.37%. This means that if our DCNN model recommends surgery for a patient, we have over 91% confidence to make this decision, with a narrow false-negative rate of mistaking TRG0 patients as non-TRG0 patients. Generally, the DCNN model was superior to the other models mentioned above; it both reduced unnecessary watchful waiting and avoided missing the optimal time window for surgery as well as controlled the false-positive rate within an acceptable range. Last but not the least, unlike MRI-based predictive models, our DCNN model can be easily applied using any standard endoscopic system. Imaging that there is an AI module linkage endoscopic equipment, when an LARC patient undergoes an endoscopy, the AI module would capture tumor lesions automatically and calculate the probabilities of TRG0 in real time. A result of TRG0 would receive a W&W strategy recommendation if the lymph nodes are radiologically negative in the dialog box, otherwise a surgery advice. Sometimes we may find residual lesions in the intestinal lumen even it shows TRG0, and a transanal excision or endoscopic excision is warranted.

Several limitations of the present study should be noted. First, this study only assessed the local luminal tumor regression grade, and did not analyze lymph node involvement or distant metastasis status. However, it has been reported that positive lymph nodes are seldom found among patients with TRG0 (Debove et al., 2016). A multimodality DCNN model comprising MRI, endoscopy images, and clinicopathological characteristics is expected to overcome this limitation. Second, as this was a retrospective study, selection bias could not be avoided; however, the present study has the largest sample size among related studies. We included a total of 1,000 patients in our study, and performed image augmentation to further expand the sample quantity. Last, our model was based on a single-center study, and has not been validated by other centers, so our results should be interpreted with caution. Despite these limitations, a DCNN algorithm based on colonoscopy images could accurately reflect heterogeneity within the tumor, and the model was not affected by population distribution, making it possible to include more patients to improve accuracy. Further prospective multi-center research studies may improve the performance of our model, and efforts to raise model interpretability (e.g., by incorporating a visual representation of the network's output) might help to increase trust in deep learning models.

In conclusion, the proposed DCNN model achieved high accuracy, sensitivity, specificity, and NPV in predicting TRG0 in LARC patients after NT, with a better performance than experienced endoscopists. This tool may serve as an ideal alternative method for monitoring treatment response during NT and could add value in identifying surgery candidates.

DATA AVAILABILITY STATEMENT

The endoscopic images and clinicopathological data are not publicly available for patient privacy protection purposes but are available from the corresponding authors (LL, JH and PL) with a signed data access agreement. The deep convolutional neural network model was developed and improved based on the standard library in open-source platforms. The improved part was saved in our institute and can be obtained from our institute if necessary. Custom codes for the deployment of the model are available https://github.com/zhanghang1989/ResNeSt.

ETHICS STATEMENT

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

Conception and design: LL, JH, and PL; analysis and interpretation of the data: XC, JC, XH, LX, YL, WL, DL and YF; drafting of the article: XC, JC and XH; critical revision of the article for important intellectual content: XH, LL, JH and PL; final approval of the article: all the authors.

FUNDING

This work was supported by the national Key R&D Program of China (No. 2017YFC1308800); National Natural Science Foundation of China (No. 81970482); The Sixth Affiliated Hospital of Sun Yat-Sen University Clinical Research-'1010' Program (No. 1010PY (2020)-63), and National Key Clinical Discipline.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2022.880981/full#supplementary-material

REFERENCES

Al-Sukhni, E., Attwood, K., Mattson, D. M., Gabriel, E., and Nurkin, S. J. (2016). Predictors of Pathologic Complete Response Following Neoadjuvant Chemoradiotherapy for Rectal Cancer. *Ann. Surg. Oncol.* 23 (4), 1177–1186. doi:10.1245/s10434-015-5017-y

Asoglu, O., Tokmak, H., Bakir, B., Demir, G., Ozyar, E., Atalar, B., et al. (2020). The Impact of Total Neo-Adjuvant Treatment on Nonoperative Management in Patients with Locally Advanced Rectal Cancer: The Evaluation of 66 Cases. *Eur. J. Surg. Oncol.* 46 (3), 402–409. doi:10.1016/j.ejso.2019.07.012

Borowski, D. W., Bradburn, D. M., Mills, S. J., Bharathan, B., Wilson, R. G., Ratcliffe, A. A., et al. (2010). Volume-Outcome Analysis of Colorectal Cancer-Related Outcomes. *Br. J. Surg.* 97 (9), 1416–1430. doi:10.1002/bjs.7111

Coraglio, M. F., Eleta, M. A., Kujaruk, M. R., Oviedo, J. H., Roca, E. L., Masciangioli, G. A., et al. (2020). Analysis of Long-Term Oncological Results of Clinical Versus Pathological Responses after Neoadjuvant Treatment in Locally Advanced Rectal Cancer. *World J. Surg. Onc* 18 (1), 313. doi:10.1186/s12957-020-02094-1

Dattani, M., Heald, R. J., Goussous, G., Broadhurst, J., São Julião, G. P., Habr-Gama, A., et al. (2018). Oncological and Survival Outcomes in Watch and Wait Patients with a Clinical Complete Response after Neoadjuvant Chemoradiotherapy for Rectal Cancer: A Systematic Review and Pooled Analysis. *Ann. Surg.* 268 (6), 955–967. doi:10.1097/sla.0000000000002761

Debove, C., Guedj, N., Tribillon, E., Maggiori, L., Zappa, M., and Panis, Y. (2016). Local Excision of Low Rectal Cancer Treated by Chemoradiotherapy: Is it Safe for All Patients with Suspicion of Complete Tumor Response? *Int. J. Colorectal Dis.* 31 (4), 853–860. doi:10.1007/s00384-016-2546-1

Glynne-Jones, R., Wyrwicz, L., Tiret, E., Brown, G., Rödel, C., Cervantes, A., et al. (2017). Rectal Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* 28 (Suppl. l_4), iv22–iv40. doi:10.1093/annonc/mdx224

Guillem, J. G., Chessin, D. B., Shia, J., Moore, H. G., Mazumdar, M., Bernard, B., et al. (2005). Clinical Examination Following Preoperative Chemoradiation for Rectal Cancer Is Not a Reliable Surrogate End point. *J. Clin. Oncol.* 23 (15), 3475–3479. doi:10.1200/jco.2005.06.114

Hiotis, S. P., Weber, S. M., Cohen, A. M., Minsky, B. D., Paty, P. B., Guillem, J. G., et al. (2002). Assessing the Predictive Value of Clinical Complete Response to Neoadjuvant Therapy for Rectal Cancer: An Analysis of 488 Patients. *J. Am. Coll. Surgeons* 194 (2), 131–135. doi:10.1016/s1072-7515(01)01159-0

Hupkens, B. J. P., Martens, M. H., Stoot, J. H., Berbee, M., Melenhorst, J., Beets-Tan, R. G., et al. (2017). Quality of Life in Rectal Cancer Patients after Chemoradiation: Watch-And-Wait Policy Versus Standard Resection - A Matched-Controlled Study. *Dis. Colon Rectum* 60 (10), 1032–1040. doi:10.1097/dcr.0000000000000862

Jiang, G., Wei, J., Xu, Y., He, Z., Zeng, H., Wu, J., et al. (2021). Synthesis of Mammogram from Digital Breast Tomosynthesis Using Deep Convolutional Neural Network with Gradient Guided cGANs. *IEEE Trans. Med. Imaging* 40 (8), 2080–2091. doi:10.1109/tmi.2021.3071544

Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., et al. (2018). Diagnosis of Thyroid Cancer Using Deep Convolutional Neural Network Models Applied to Sonographic Images: A Retrospective, Multicohort, Diagnostic Study. *Lancet Oncol.* 20, 193–201. doi:10.1016/S1470-2045(18)30762-9

Liu, S., Zhong, G.-X., Zhou, W.-X., Xue, H.-D., Pan, W.-D., Xu, L., et al. (2018). Can Endorectal Ultrasound, MRI, and Mucosa Integrity Accurately Predict the Complete Response for Mid-Low Rectal Cancer after Preoperative Chemoradiation? A Prospective Observational Study from a Single Medical Center. *Dis. Colon Rectum* 61 (8), 903–910. doi:10.1097/dcr.0000000000001135

Maas, M., Nelemans, P. J., Valentini, V., Das, P., Rödel, C., Kuo, L.-J., et al. (2010). Long-Term Outcome in Patients with a Pathological Complete Response after Chemoradiation for Rectal Cancer: A Pooled Analysis of Individual Patient Data. *Lancet Oncol.* 11 (9), 835–844. doi:10.1016/s1470-2045(10)70172-8

Marijnen, C. A. M. (2015). Organ Preservation in Rectal Cancer: Have All Questions Been Answered? *Lancet Oncol.* 16 (1), e13–e22. doi:10.1016/s1470-2045(14)70398-5

Ortholan, C., Romestaing, P., Chapet, O., and Gerard, J. P. (2012). Correlation in Rectal Cancer between Clinical Tumor Response after Neoadjuvant Radiotherapy and Sphincter or Organ Preservation: 10-year Results of the Lyon R 96-02 Randomized Trial. *Int. J. Radiat. Oncol. Biol. Phys.* 83 (2), e165–e171. doi:10.1016/j.ijrobp.2011.12.002

Pang, K., Rao, Q., Qin, S., Jin, L., Yao, H., and Zhang, Z. (2019). Prognosis Comparison between Wait and Watch and Surgical Strategy on Rectal Cancer Patients after Treatment with Neoadjuvant Chemoradiotherapy: A Meta-Analysis. *Therap Adv. Gastroenterol.* 12, 1756284819892477. doi:10.1177/1756284819892477

Pinto, J. C., Pereira, A. D., Pimenta, A., Pedro, C., Fernandez, G., Marques, I., et al. (2020). Low Rectal Cancer Treatment Strategies: A Cohort Study Assessing Watch and Wait. *J. Cancer Res. Clin. Oncol.* 146 (10), 2631–2638. doi:10.1007/s00432-020-03248-0

Roh, M. S., Colangelo, L. H., O'Connell, M. J., Yothers, G., Deutsch, M., Allegra, C. J., et al. (2009). Preoperative Multimodality Therapy Improves Disease-free Survival in Patients with Carcinoma of the Rectum: NSABP R-03. *J. Clin. Oncol.* 27 (31), 5124–5130. doi:10.1200/jco.2009.22.0467

Sauer, R., Becker, H., Hohenberger, W., Rödel, C., Wittekind, C., Fietkau, R., et al. (2004). Preoperative versus Postoperative Chemoradiotherapy for Rectal Cancer. *N. Engl. J. Med.* 351 (17), 1731–1740. doi:10.1056/nejmoa040694

Sebag-Montefiore, D., Stephens, R. J., Steele, R., Monson, J., Grieve, R., Khanna, S., et al. (2009). Preoperative Radiotherapy Versus Selective Postoperative Chemoradiotherapy in Patients with Rectal Cancer (MRC CR07 and NCIC-CTG C016): A Multicentre, Randomised Trial. *Lancet* 373 (9666), 811–820. doi:10.1016/s0140-6736(09)60484-0

Simpson, G., Hopley, P., Wilson, J., Day, N., Haworth, A., Montazeri, A., et al. (2020). Long-Term Outcomes of Real World 'Watch and Wait' Data for Rectal Cancer after Neoadjuvant Chemoradiotherapy. *Colorectal Dis.* 22 (11), 1568–1576. doi:10.1111/codi.15177

Tang, Z., Zhang, X.-Y., Liu, Z., Li, X.-T., Shi, Y.-J., Wang, S., et al. (2019). Quantitative Analysis of Diffusion Weighted Imaging to Predict

Pathological Good Response to Neoadjuvant Chemoradiation for Locally Advanced Rectal Cancer. *Radiother. Oncol.* 132, 100–108. doi:10.1016/j.radonc.2018.11.007

Valentini, V., van Stiphout, R. G. P. M., Lammering, G., Gambacorta, M. A., Barba, M. C., Bebenek, M., et al. (2011). Nomograms for Predicting Local Recurrence, Distant Metastases, and Overall Survival for Patients with Locally Advanced Rectal Cancer on the Basis of European Randomized Clinical Trials. *J. Clin. Oncol.* 29 (23), 3163–3172. doi:10.1200/jco.2010.33.1595

Valentini, V., van Stiphout, R. G. P. M., Lammering, G., Gambacorta, M. A., Barba, M. C., Bebenek, M., et al. (2015). Selection of Appropriate End-Points (pCR vs 2yDFS) for Tailoring Treatments with Prediction Models in Locally Advanced Rectal Cancer. *Radiother. Oncol.* 114 (3), 302–309. doi:10.1016/j.radonc.2015.02.001

van der Sande, M. E., Maas, M., Melenhorst, J., Breukink, S. O., van Leerdam, M. E., and Beets, G. L. (2021). Predictive Value of Endoscopic Features for a Complete Response after Chemoradiotherapy for Rectal Cancer. *Ann. Surg.* 274 (6), e541–e547. doi:10.1097/sla.0000000000003718

van der Valk, M. J. M., Hilling, D. E., Bastiaannet, E., Meershoek-Klein Kranenbarg, E., Beets, G. L., Figueiredo, N. L., et al. (2018). Long-Term Outcomes of Clinical Complete Responders after Neoadjuvant Treatment for Rectal Cancer in the International Watch & Wait Database (IWWD): An International Multicentre Registry Study. *Lancet* 391 (10139), 2537–2545. doi:10.1016/S0140-6736(18)31078-X

Weiser, M. R. (2018). AJCC 8th Edition: Colorectal Cancer. *Ann. Surg. Oncol.* 25 (6), 1454–1455. doi:10.1245/s10434-018-6462-1

Xie, Y., Zhang, J., Xia, Y., and Shen, C. (2020). A Mutual Bootstrapping Model for Automated Skin Lesion Segmentation and Classification. *IEEE Trans. Med. Imaging* 39 (7), 2482–2493. doi:10.1109/tmi.2020.2972964

Zhang, X.-Y., Wang, L., Zhu, H.-T., Li, Z.-W., Ye, M., Li, X.-T., et al. (2020). Predicting Rectal Cancer Response to Neoadjuvant Chemoradiotherapy Using Deep Learning of Diffusion Kurtosis MRI. *Radiology* 296 (1), 56–64. doi:10.1148/radiol.2020190936

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, Chen, He, Xu, Liu, Lin, Luo, Feng, Lian, Hu and Lan. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 04 May 2022
doi: 10.3389/fmolb.2022.880291

Predicting the Stone-Free Status of Percutaneous Nephrolithotomy With the Machine Learning System: Comparative Analysis With Guy's Stone Score and the S.T.O.N.E Score System

Hong Zhao[1], *Wanling Li*[2], *Junsheng Li*[1], *Li Li*[1], *Hang Wang*[2] *and Jianming Guo*[2]*

[1]Shanghai Xuhui Central Hospital, Shanghai, China, [2]Zhongshan Hospital, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Mohammed Shahait,
King Hussein Medical Center, Jordan
Yasser A. Noureldin,
Benha University, Egypt

***Correspondence:**
Jianming Guo
drguojm@126.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 21 February 2022
Accepted: 07 April 2022
Published: 04 May 2022

Citation:
Zhao H, Li W, Li J, Li L, Wang H and Guo J (2022) Predicting the Stone-Free Status of Percutaneous Nephrolithotomy With the Machine Learning System: Comparative Analysis With Guy's Stone Score and the S.T.O.N.E Score System. Front. Pharmacol. 9:880291. doi: 10.3389/fmolb.2022.880291

Purpose: The aim of the study was to use machine learning methods (MLMs) to predict the stone-free status after percutaneous nephrolithotomy (PCNL). We compared the performance of this system with Guy's stone score and the S.T.O.N.E score system.

Materials and Methods: Data from 222 patients (90 females, 41%) who underwent PCNL at our center were used. Twenty-six parameters, including individual variables, renal and stone factors, and surgical factors were used as input data for MLMs. We evaluated the efficacy of four different techniques: Lasso-logistic (LL), random forest (RF), support vector machine (SVM), and Naive Bayes. The model performance was evaluated using the area under the curve (AUC) and compared with that of Guy's stone score and the S.T.O.N.E score system.

Results: The overall stone-free rate was 50% (111/222). To predict the stone-free status, all receiver operating characteristic curves of the four MLMs were above the curve for Guy's stone score. The AUCs of LL, RF, SVM, and Naive Bayes were 0.879, 0.803, 0.818, and 0.803, respectively. These values were higher than the AUC of Guy's score system, 0.800. The accuracies of the MLMs (0.803% to 0.818%) were also superior to the S.T.O.N.E score system (0.788%). Among the MLMs, Lasso-logistic showed the most favorable AUC.

Conclusion: Machine learning methods can predict the stone-free rate with AUCs not inferior to those of Guy's stone score and the S.T.O.N.E score system.

Keywords: machine learning, prediction, percutaneous nephrolithotomy, stone-free status, Guy's stone score, S.T.O.N.E score system

Abbreviations: MLMs, machine learning methods; PCNL, percutaneous nephrolithotomy; LL, Lasso logistic; RF, random forest; SVM, support vector machine; AUC, area under the curve; GSS, Guy's score system; CROES, Clinical Research Office of the Endourological Society; SFR, stone-free rate; NCCT, non-contrast CT; EBL, estimated blood loss; LOT, longer operative time; LOS, length of stay

INTRODUCTION

Since the first description of the technique in 1976 (Fernstro¨m and Johansson, 1976), percutaneous nephrolithotomy has been widespread for the treatment of renal calculi. It is the golden standard for the treatment of 2-cm kidney stones (Miernik et al., 2014). PCNL's success rate is between 56% and 96% in various series (Matlaga et al., 2005; Akman et al., 2011; Rosette et al., 2011; Labadie et al., 2015). Many factors contribute to the success of stone clearance including the stone size, location, number, and grade of hydronephrosis, as well as surgeon's experience. To predict the outcomes after PCNL, several scoring systems have been devised including Guy's stone score, S.T.O.N.E nephrolithometry system, CROES nephrolithometry nomogram, and S-ReSC score (Al Adl et al., 2020). Guy's stone score is easy to apply and has been validated in multiple studies. The S.T.O.N.E. score is based on factors determined through CT imaging, which is the currently preferred imaging modality for patients with nephrolithiasis (Noureldin et al., 2015a). The CROES nomogram was developed from data in a large multicenter database and has high statistical power. Determination of the S-ReSC score relies on stone location only, providing a simple approach to grading disease complexity (Noureldin et al., 2015b). Each system has advantages and disadvantages, but several studies suggest that their ability to predict the stone-free rate is comparable (Wu and Okeke, 2017).

Machine learning techniques have been used extensively in the field of clinical medicine, especially when used for the construction of prediction models. The outperformance of ML over conventional data analysis models has been shown in the urology-oncology literature (Hung et al., 2018; Andras et al., 2020; Rodrigo. et al., 2020; Ström et al., 2020).

In predicting post-lithotripsy outcomes with machine learning, there are only three studies published until now (De Perrot et al., 2019; Tayyebe. et al., 2019; Aminsharifi et al., 2020). Aminsharifi et al. (2020) first used the machine learning method for predicting post-PCNL outcomes compared to current scoring systems. They found machine learning-based software was superior in predicting SFS after PCNL, with an AUC of 0.915 compared to 0.615 (GSS) and 0.621 (CROES nomograms) ($p < 0.01$). More than 20 variables of 146 patients were inputted for the training of machine learning in their study. Alireza used a support vector machine (SVM) as the machine learning technique. We know that the machine learning algorithm includes some other methods, such as decision trees, random forests, artificial neural networks, Bayesian learning, Deep Learning, and so on. In this study, we used four machine learning methods (Lasso logistic, random forests, SVM, and Naive Bayes) to predict the SFS of PCNL with the information of 222 patients. We compared the outperformance of ML to Guy's score and the S.T.O.N.E score system at the same time.

Patients and Methods

The study was approved by the independent ethics committee of Xu-hui Central Hospital. Between July 2017 and January 2020, 222 patients who underwent PCNL performed by one single surgeon (Dr. G.J.M.) were included in this retrospective study. All patients had computed tomography (CT) scans and IVP before surgery. Normal preoperative coagulation and negative urine cultures were verified.

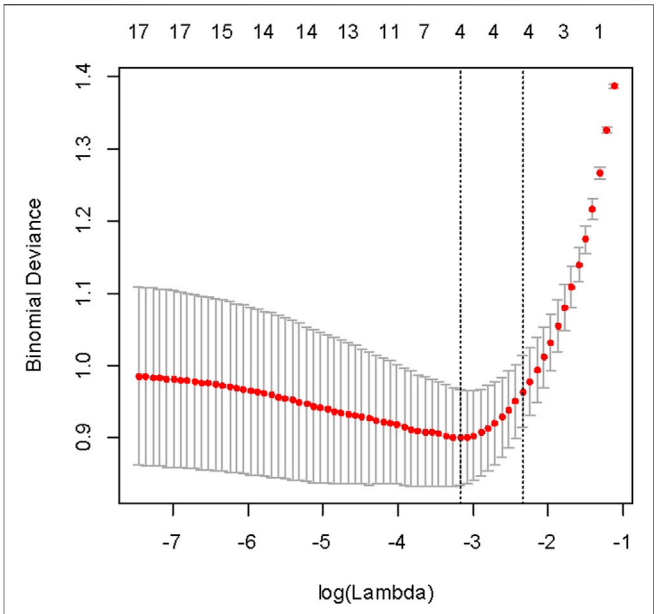


FIGURE 1 | Selecting lambda to screen characteristic variables.

All percutaneous accesses were performed under general anesthesia and in a prone position after retrograde ureteral catheterization. Access to the selected calyx was performed by Dr. G.J.M with the aid of ultrasound guidance by using an 18-gauge needle. The tract was dilated with serial dilators from 8F to 20F sheath. An 18F nephroscope (Wolf) was used to inspect the sheath, and we used a holmium laser to fragment stones with the power ranging from 60 to 90 W. Every case was demanded to place an internal ureteral stent on a suspect for the presence of mobile residual stones. A 14F nephrostomy tube was placed in the renal pelvis or the involved calyx for most patients.

Antibiotic prophylaxis was used with the second-generation cephalosporin. The medication was completed after the nephrostomy tube was removed.

Plain radiography of the kidneys, ureters, and bladders was obtained from postoperative day 1 to day 3, according to the state of the patient. The nephrostomy tube was removed when there were neither stone residues nor clinically insignificant residual fragments (diameter less than 4 mm). (Harraz et al., 2017).

All patients were asked to take out the stent for outpatient service 1 or 2 months after the surgery. If there were residual stones, they would have repeated PCNL, ureteroscopy, and shock wave lithotripsy (SWL). After that, all patients were evaluated with an ultrasound test or non-contrast CT scan after 3–6 months postoperatively. All patients accepted follow-ups for at least 1 year. PCNL was considered successful when the patient was stone-free or did not need any further intervention [clinically insignificant residual stone fragments (CIRF)] (Rassweiler et al., 2000).

TABLE 1 | Preoperative factors include individual variables and renal and stone factors.

Age (mean ± SD) (years)	54.81 ± 13.31	%
Gender (male/female)	132/90	59.46
Guy's score	3.27 ± 0.87	
S.T.O.N.E score	8.91 ± 1.82	
Stone burden (mm2)[a]	563.4 ± 517.6	
History of diabetes n (%)	45	20.27
History of hypertension n (%)	70	31.53
History of hyperlipidemia n (%)	39	17.57
Solitary kidney n (%)	18	8.11
Renal insufficiency n (%)	30	13.51
Anemia n (%)	29	31.53
Preoperative urinary infection n (%)	111	50.00
Previous surgery in target kidney n (%)	77	34.68
SMWL	22	9.91
URSL	21	9.46
PCNL	24	10.81
Open surgery	26	11.71
Hydronephrosis n (%)	112	50.45
Stone location n (%)		
Upper calyx	116	52.25
Mid calyx	136	61.26
Lower calyx	164	73.87
Renal pelvis	160	72.07
Ureter	50	22.52

[a]*Stone burden = Length × Width × 0.78.*

TABLE 2 | Postoperative outcome variable (*n* = 222).

Hospitalization day	11.15 ± 4.98	10.49 (%)
Transfusion n (%)	9	4.1
Fever n (%)	42	18.9
Septicemia n (%)	19	8.6
Interventional therapy n (%)	3	1.4
Pleural injury n (%)	2	0.9
Ancillary procedures n (%)	12	5.4
Stone-free rate n (%)	111	50.0

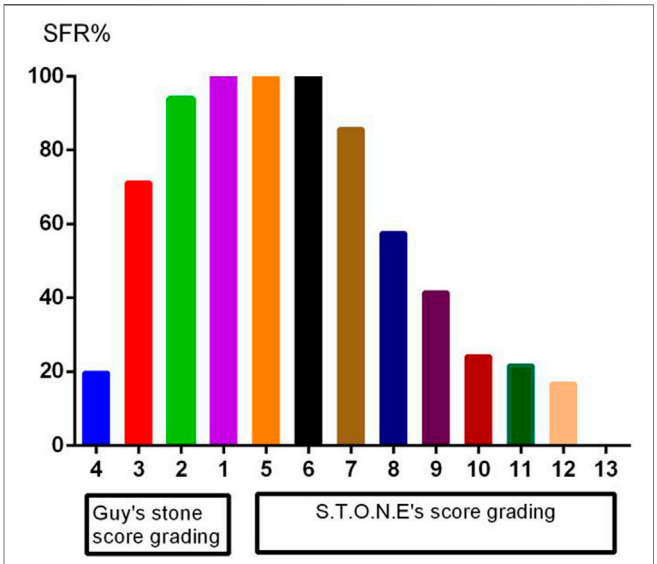


FIGURE 2 | The stone-free rate in each subgroup of GSS grades and the S.T.ON.E score systems.

Machine Learning Methods

Four types of supervised machine learning algorithms (Lasso logistic, random forests, SVM, and Naive Bayes) were applied in this study. A set of input variables comprising individual variables (age, sex, hypertension, diabetes, hyperlipidemia, urinary infection, renal insufficiency, preoperative hemoglobin, use of anticoagulants or antiplatelet medications, renal and stone factors (previous surgery, stone burden, stone location, and hydronephrosis), surgical factors (postoperative fever, septicemia, need for transfusion, length of stay, stone-free status, and ancillary procedures)) were included. The results of the stone-free status were entered as binary values: 1 (stone residues) and 0 (clinically insignificant residual stone fragments).

The machine learning models were fitted using scikit-learn 0.18 modules of Python throughout this study. Using lasso regularization and cross-validation (*n* fold = 10) to select the best regression, we selected lambda with 1se.lambda to screen characteristic variables. The selected variables include stone size, stone location (top/middle/bottom), and a total of four variables (**Figure 1**).

The original data set is randomly divided into the training set and the test set at 7:3 (156: 66). Lasso-logistic, SVM, and Naive Bayes considered the results of lasso regression screening as independent variables to establish a model and calculate the prediction accuracy.

The RF model is a machine learning model built on decision trees. In the decision tree, each node of the tree splits the data into two groups using a cutoff value within one of the features. The RF method can minimize the effect of the overfitting problem by creating an ensemble of randomized decision trees, each of which overfits the data and averages the results to find a better classification.

Statistical Analysis

Continuous variables were compared using the independent sample Student's t-test. The model performance was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), which provides a measure of the discriminatory performance of the model. Sensitivity is the proportion of true positives that are classified as such; specificity measures the proportion of correctly identified true negatives; and accuracy is the proportion of correct predictions.

RESULTS

A total of 222 patients (132 males, 59.5%) were enrolled. The mean age was 54.8 ± 13.3 years, and the mean stone burden was 563.4 ± 517.6 mm^2. The mean Guy's score was 3.2 ± 0.9, and the mean S.T.O.N.E. score was 8.9 ± 1.8. **Table 1** shows the preoperative factors including individual variables and renal and stone factors. **Table 2** shows the actual postoperative data for these patients. The overall SFS was 50% (111/222). **Figure 2**

TABLE 3 | AUC, sensitivity, specificity, and accuracy of each prediction method for the results of the stone-free status.

Outcome	Lasso logistic	Random forest	Support vector machine	Naive Bayes	Guy's score	S.T.O.N.E score system
AUC	0.879	0.803	0.818	0.803	0.800	0.844
Sensitivity (%)	0.7576	0.7576	0.7576	0.8333	0.8180	0.7575
Specificity (%)	0.8788	0.8485	0.8788	0.7778	0.8480	0.8181
Accuracy (%)	0.8181	0.8030	0.8182	0.8030	0.8333	0.7878

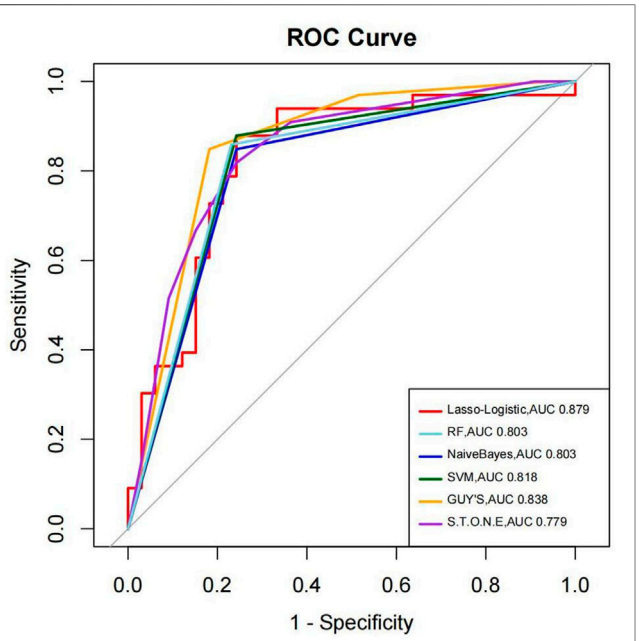


FIGURE 3 | The ROC curves of the four MLMs as well as the GSS and S.T.O.N.E score system.

shows the stone-free rate in each subgroup of GSS grades and the S.T.O.N.E score systems. The number of fever and infections during hospitalization was 18.9% (42) and 8.6% (19). Postoperative blood transfusion due to significant blood loss happened in nine patients (4.1%). With the follow-ups for at least 1 year, there were 12 patients (5.4%) who accepted ancillary procedures to manage residual renal stones.

We have used four machine learning methods to analyze the outcomes to predict the stone-free status. **Table 3** shows the AUC, sensitivity, specificity, and accuracy of each prediction method to the results of the stone-free status. When using AUC as a measure of the predictive model performance, as shown in **Table 3**, the AUC of Lasso logistic was 0.879. It was superior to those of RF, SVM, and Naive Bayes (0.803, 0.818, and 0.803, respectively). The AUCs of the GSS and S.T.O.N.E were 0.800 and 0.844, respectively, which were lower than the Lasso logistic. **Figure 3** shows the ROC curves of the four MLMs, as well as the GSS and S.T.O.N.E score system.

As shown in **Table 3**, the accuracies of the four MLMs were also superior to those of the S.T.O.N.E score system. The sensitivities of the MLMs were 75.8–83.3%, which were higher than the S.T.O.N.E. score system. The machine learning system of LL recognized stone burden and stone location as the most highly weighted preoperative factors affecting the post-PCNL-SFR.

DISCUSSION

The incidence and prevalence of kidney stones have increased by three times over the past 4 decades (Thongprayoon et al., 2020). The prevalence of kidney stones is estimated at about 5–10% in Europe, 4% in South America, and 1–19% in Asia currently (Sorokin et al., 2017; Liu et al., 2018). Without a doubt, kidney stones represent a considerable burden for public healthcare systems.

Thomas et al. (2011) were the first to introduce Guy's stone score (GSS) to predict the success of the post-PCNL stone-free status (SFS). The model is reproducible, provides quick and easy office-based categorization of renal stones in four grades based on stone shape and configuration, and correlates well with the SFS; however, it fails to take into account the size and density of the stone. The S.T.O.N.E. nephrolithometry scoring system of Okhunov et al. (Zhamshid. et al., 2013) is based on non-contrast CT (NCCT) having five variables; a score of 5–6 (low complexity) has an overall SFS of 94–100%, and a score 9–13 (high complexity) has an overall SFS of 27–64%. Also, greater S.T.O.N.E. scores are associated with a greater estimated blood loss (EBL), longer operative times (LOTs), and increased length of stay (LOS) in hospital. Smith et al. (2013) developed the CROES (Clinical Research Office of the Endourological Society) nomogram to predict the SFS after PCNL based on a global database study of 5,830 patients. Six characteristics (stone burden, number, location, multiple, staghorn, and institute-level case volume) are included in this nomogram. It achieved a remarkable prediction accuracy of 76%, but it is laborsome and time-consuming.

Many studies have compared the predictive performance of these score systems in post-PCNL SFR. Most studies have examined the performance of these scoring systems to predict SFR equally but not equally to predict complications. The AUC ranges from 0.63 to 0.853 (Wu and Okeke, 2017), and the different scoring system has its drawbacks or limitations. For example, in Guy's score system, partial staghorn stone was not clearly defined. The S.T.O.N.E. nephrolithometry scoring system relies solely on preoperative CT. The CROES nomogram requires information that might not be readily available (case volume and treatment history). So one simpler and easier application stone score system is needed nowadays. Alireza and his colleagues (Aminsharifi et al., 2017) were the first to use machine learning

methods to evaluate the stone-free rate and complications after PCNL. They used ANN to predict the stone-free rate. The accuracy was 81.0–98.2%. The AUC was 0.861. In 2019, his team (Aminsharifi et al., 2020) reported they used software to predict the SFR after PCNL with the AUC of 0.915. In our study, we used four machine learning methods to predict the SFR of PCNL compared with Guy's system and the S.T.O.N.E. nephrolithometry system. The machine learning methods (MLMs) include Lasso logistic, random forests, SVM, and Naive Bayes. The AUC of the MLMs was superior than that of Guy's stone score system. The sensitivity and accuracy of MLMs were superior to that of the S.T.O.N.E. nephrolithometry system.

Machine learning is built on the statistical framework. Different approaches are designed to make the most accurate prediction possible. It has been proved to have a good performance to predict the SFR post-PCNL. Although we did not have an advantageous performance of AUC of 0.915 (Aminsharifi et al., 2020), in this study, we found the MLMs could predict the stone-free rate with the AUC not inferior to that of Guy's stone score or the S.T.O.N.E score system. The machine learning algorithm mainly includes random forests, decision trees, artificial neural networks, Bayesian learning, and Deep learning. Each approach has its advantage and disadvantage. We have tried four methods to predict the stone-free rate in this study, and all of them got a fairly superior performance, as well as the clinical scoring systems being currently available. Machine learning methods are a good tool to predict the stone-free rate with AUCs after PCNL.

So far, in the field of urinary stones, there have been few studies using machine learning methods to predict operative outcomes or help make operative decisions. As one author commented (Peng et al., 2021), to improve the application of MLMs in uritholiasis, two categories should be considered: first, more people including urologists, statisticians, and computer experts need to be involved in this project; second, more data from different regions or population should be collected for future event prediction. We need to establish, manage, and share a cross-country or nationwide database, through which machine learning or AI would contribute to the field of calculi or other issues in the near future.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

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: protocol and writing of the manuscript. WL: data analysis. JL: data collection. LL: management. JG: project development (corresponding author). HZ and WL contributed equally to this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.880291/full#supplementary-material

REFERENCES

Akman, T., Binbay, M., Yuruk, E., Sari, E., Seyrek, M., Kaba, M., et al. (2011). Tubeless Procedure Is Most Important Factor in Reducing Length of Hospitalization after Percutaneous Nephrolithotomy: Results of Univariable and Multivariable Models. *Urology* 77 (2), 299–304. doi:10.1016/j.urology.2010.06.060

Al Adl, A. M., Mohey, A., Abdel Aal, A., Abu-Elnasr, H. A. F., El Karamany, T., Noureldin, Y. A., et al. (2020). Percutaneous Nephrolithotomy Outcomes Based on S.T.O.N.E., GUY, CROES, and S-ReSC Scoring Systems: The First Prospective Study. *J. Endourology* 34, 1223–1228. doi:10.1089/end.2019.0856

Aminsharifi, A., Irani, D., Pooyesh, S., Parvin, H., Dehghani, S., Yousofi, K., et al. (2017). Artificial Neural Network System to Predict the Postoperative Outcome of Percutaneous Nephrolithotomy. *J. Endourology* 31 (5), 461–467. doi:10.1089/end.2016.0791

Aminsharifi, A., Irani, D., Tayebi, S., Jafari Kafash, T., Shabanian, T., and Parsaei, H. (2020). Predicting the Postoperative Outcome of Percutaneous Nephrolithotomy with Machine Learning System: Software Validation and Comparative Analysis with Guy's Stone Score and the CROES Nomogram. *J. Endourology* 34 (6), 692–699. doi:10.1089/end.2019.0475

Andras, I., Mazzone, E., van Leeuwen, F. W. B., De Naeyer, G., van Oosterom, M. N., Beato, S., et al. (2020). Artificial Intelligence and Robotics: a Combination that Is Changing the Operating Room. *World J. Urol.* 38 (10), 2359–2366. doi:10.1007/s00345-019-03037-6

De Perrot, T., Hofmeister, J., Burgermeister, S., Martin, S. P., Feutry, G., Klein, J., et al. (2019). Differentiating Kidney Stones from Phleboliths in Unenhanced Low-Dose Computed Tomography Using Radiomics and Machine Learning. *Eur. Radiol.* 29 (9), 4776–4782. doi:10.1007/s00330-019-6004-7,

Fernstro¨m, I., and Johansson, B. (1976). Percutaneous Pyelolithotomy. A New Extraction Technique. *Scand. J. Urol. Nephrol.* 10, 257–259.

Harraz, A. M., Osman, Y., El-Nahas, A. R., Elsawy, A. A., Fakhreldin, I., Mahmoud, O., et al. (2017). Residual Stones after Percutaneous Nephrolithotomy: Comparison of Intraoperative Assessment and Postoperative Non-contrast Computerized Tomography. *World J. Urol.* 35 (8), 1241–1246. doi:10.1007/s00345-016-1990-4

Hung, A. J., Chen, J., Che, Z., Nilanon, T., Jarc, A., Titus, M., et al. (2018). Utilizing Machine Learning and Automated Performance Metrics to Evaluate Robot-Assisted Radical Prostatectomy Performance and Predict Outcomes. *J. Endourology* 32 (5), 438–444. doi:10.1089/end.2018.0035

Labadie, K., Okhunov, Z., Akhavein, A., Moreira, D. M., Moreno-Palacios, J., Del Junco, M., et al. (2015). Evaluation and Comparison of Urolithiasis Scoring Systems Used in Percutaneous Kidney Stone Surgery. *J. Urol.* 193 (1), 154–159. doi:10.1016/j.juro.2014.07.104

Liu, Y., Chen, Y., Liao, B., Luo, D., Wang, K., Li, H., et al. (2018). Epidemiology of Urolithiasis in Asia. *Asian J. Urol.* 5 (4), 205–214. doi:10.1016/j.ajur.2018.08.007

Matlaga, B. R., Kim, S. C., and Lingeman, J. E. (2005). Improving Outcomes of Percutaneous Nephrolithotomy: Access. *EAU Update Ser.* 3, 37–43. doi:10.1016/j.euus.2004.11.002

Miernik, A., Schoenthaler, M., Wilhelm, K., Wetterauer, U., Zyczkowski, M., Paradysz, A., et al. (2014). Combined Semirigid and Flexible Ureterorenoscopy via a Large Ureteral Access Sheath for Kidney Stones >2 Cm: a Bicentric Prospective Assessment. *World J. Urol.* 32 (3), 697–702. doi:10.1007/s00345-013-1126-z

Noureldin, Y. A., Elkoushy, M. A., and Andonian, S. (2015). Which Is Better? Guy's versus S.T.O.N.E. Nephrolithometry Scoring Systems in Predicting Stone-free Status post-percutaneous Nephrolithotomy. *World J. Urol.* 33, 1821–1825. doi:10.1007/s00345-015-1508-5

Noureldin, Y., Elkoushy, M., and Andonian, S. (2015). External Validation of the S.T.O.N.E. Nephrolithometry Scoring System. *Cuaj* 9, 190–195. doi:10.5489/cuaj.2652

Peng, L., Pan, J., Yang, D., and Zhong, W. (2021). The prospect of Machine Learning in Predicting post-lithotripsy Outcomes. *World J. Urol.* 39 (11), 4287–4288. doi:10.1007/s00345-020-03377-8

Rassweiler, J. J., Renner, C., and Eisenberger, F. (2000). The Management of Complex Renal Stones. *BJU Int.* 86, 919–928. doi:10.1046/j.1464-410x.2000.00906.x

Rodrigo., S-I., Hein, S., Reis, G., Christian., G., and Miernik, A. (2020). Current and Future Applications of Machine and Deep Learning in Urology: a Review of the Literature on Urolithiasis, Renal Cell Carcinoma, and Bladder and Prostate Cancer. *World J. Urol.* 38 (10), 2329–2347.

Rosette, J. d. l., Assimos, D., Desai, M., Gutierrez, J., Lingeman, J., Scarpa, R., et al.CROES PCNL Study Group (2011). The Clinical Research Office of the Endourological Society Percutaneous Nephrolithotomy Global Study: Indications, Complications, and Outcomes in 5803 Patients. *J. Endourology* 25 (1), 11–17. doi:10.1089/end.2010.0424

Smith, A., Averch, T. D., Shahrour, K., Opondo, D., Daels, F. P. J., Labate, G., et al.CROES PCNL Study Group (2013). A Nephrolithometric Nomogram to Predict Treatment success of Percutaneous Nephrolithotomy. *J. Urol.* 190 (1), 149–156. doi:10.1016/j.juro.2013.01.047

Sorokin, I., Mamoulakis, C., Miyazawa, K., Rodgers, A., Talati, J., and Lotan, Y. (2017). Epidemiology of Stone Disease across the World. *World J. Urol.* 35 (9), 1301–1320. doi:10.1007/s00345-017-2008-6

Ström, P., Kartasalo, K., Olsson, H., Solorzano, L., Delahunt, B., Berney, D. M., et al. (2020). Artificial Intelligence for Diagnosis and Grading of Prostate Cancer in Biopsies: a Population-Based, Diagnostic Study. *Lancet Oncol.* 21 (2), 222–232. doi:10.1016/s1470-2045(19)30738-7

Tayyebe., S., Parsaei, H., Aminsharifi, A., Mehdi., M. M., Jahromi Amin, T., Pouyesh, S., et al. (2019). An Artificial Intelligence-Based Clinical Decision Support System for Large Kidney Stone Treatment. *Australas. Phys. Eng. Sci. Med.* 42 (3), 771–779.

Thomas, K., Smith, N. C., Hegarty, N., and Glass, J. M. (2011). The Guy's Stone Score-Ggrading the Complexity of Percutaneous Nephrolithotomy Procedures. *Urology* 78 (2), 277–281. doi:10.1016/j.urology.2010.12.026

Thongprayoon, C., Krambeck, A. E., and Rule, A. D. (2020). Determining the True burden of Kidney Stone Disease. *Nat. Rev. Nephrol.* 16 (12), 736–746. doi:10.1038/s41581-020-0320-7

Wu, W. J., and Okeke, Z. (2017). Current Clinical Scoring Systems of Percutaneous Nephrolithotomy Outcomes. *Nat. Rev. Urol.* 14, 459–469. doi:10.1038/nrurol.2017.71

Zhamshid., O., Friedlander Justin, I., George Arvin, K., Duty Brian, D., Moreira Daniel, M., Srinivasan Arun, K., et al. (2013). Hillelsohn JoelS.T.O.N.E. Nephrolithometry: Novel Surgical Classification System for Kidney Calculi. *Urology* 81 (6), 1154–1159.

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 Zhao, Li, Li, Li, Wang 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 | Frontiers in Genetics

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fgene.2022.880093

Multi-Size Deep Learning Based Preoperative Computed Tomography Signature for Prognosis Prediction of Colorectal Cancer

Cheng-Hang Li[1,2,3†], Du Cai[1,2†], Min-Er Zhong[1,2†], Min-Yi Lv[1,2], Ze-Ping Huang[1,2], Qiqi Zhu[4], Chuling Hu[1,2], Haoning Qi[1,2], Xiaojian Wu[1,2] and Feng Gao[1,2*]*

[1]Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2]Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Disease, Supported by National Key Clinical Discipline, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [3]School of Computer Science and Engineering, Guangzhou Higher Education Mega Center, Sun Yat-sen University, Guangzhou, China, [4]Ningbo Medical Center Lihuili Hospital, Ningbo, China

OPEN ACCESS

Edited by:
Liang Zhao,
Hubei University of Medicine, China

Reviewed by:
Guixi Zheng,
Shandong University, China
Andrea Tangherloni,
University of Bergamo, Italy

***Correspondence:**
Xiaojian Wu
wuxjian@mail.sysu.edu.cn
Feng Gao
gaof57@mail.sysu.edu.cn

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

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 22 February 2022
Accepted: 25 April 2022
Published: 12 May 2022

Citation:
Li C-H, Cai D, Zhong M-E, Lv M-Y, Huang Z-P, Zhu Q, Hu C, Qi H, Wu X and Gao F (2022) Multi-Size Deep Learning Based Preoperative Computed Tomography Signature for Prognosis Prediction of Colorectal Cancer. Front. Genet. 13:880093. doi: 10.3389/fgene.2022.880093

Background: Preoperative and postoperative evaluation of colorectal cancer (CRC) patients is crucial for subsequent treatment guidance. Our study aims to provide a timely and rapid assessment of the prognosis of CRC patients with deep learning according to non-invasive preoperative computed tomography (CT) and explore the underlying biological explanations.

Methods: A total of 808 CRC patients with preoperative CT (development cohort: n = 426, validation cohort: n = 382) were enrolled in our study. We proposed a novel end-to-end Multi-Size Convolutional Neural Network (MSCNN) to predict the risk of CRC recurrence with CT images (CT signature). The prognostic performance of CT signature was evaluated by Kaplan-Meier curve. An integrated nomogram was constructed to improve the clinical utility of CT signature by combining with other clinicopathologic factors. Further visualization and correlation analysis for CT deep features with paired gene expression profiles were performed to reveal the molecular characteristics of CRC tumors learned by MSCNN in radiographic imaging.

Results: The Kaplan-Meier analysis showed that CT signature was a significant prognostic factor for CRC disease-free survival (DFS) prediction [development cohort: hazard ratio (HR): 50.7, 95% CI: 28.4–90.6, $p < 0.001$; validation cohort: HR: 2.04, 95% CI: 1.44–2.89, $p < 0.001$]. Multivariable analysis confirmed the independence prognostic value of CT signature (development cohort: HR: 30.7, 95% CI: 19.8–69.3, $p < 0.001$; validation cohort: HR: 1.83, 95% CI: 1.19–2.83, $p = 0.006$). Dimension reduction and visualization of CT deep features demonstrated a high correlation with the prognosis of CRC patients. Functional pathway analysis further indicated that CRC patients with high CT signature presented down-regulation of several immunology pathways. Correlation analysis found that CT deep features were mainly associated with activation of metabolic and proliferative pathways.

Conclusions: Our deep learning based preoperative CT signature can effectively predict prognosis of CRC patients. Integration analysis of multi-omic data revealed that some molecular characteristics of CRC tumor can be captured by deep learning in CT images.

Keywords: deep learning, colorectal cancer, prognosis, nomogram, pathway analysis

INTRODUCTION

Colorectal cancer (CRC) is one of the most prevalent cancers and has become the third leading cause of cancer death (Siegel et al., 2020). Stratification of CRC patients is quite essential to design more accurate and personalized treatment according to their clinical characteristics (Sorbye et al., 2007). Though the current tumor-node-metastasis (TNM) system has been used for guiding treatment decisions of CRC patients for over 50 years (Nagtegaal et al., 2011), it is still inadequate for accurately assessing the prognosis of some colorectal patients, especially for patients in clinical stage II and III (Joachim et al., 2019). Even with the same clinical stage, patients may be suitable to different treatment options before and after surgery as heterogeneity of CRC (Molinari et al., 2018). Thus, prognostic analysis of CRC patients and evaluation of their preoperative and postoperative interventional treatment options are recent research hotspots.

Previous studies on the molecular basis of cancer and the discovery of cancer associated genes, oncogenes and tumor suppressor genes indicates that cancer is a genetic disease (Pierotti, 2017), which determines a natural advantage for cancer survival analysis with genomics data (Walther et al., 2009; Yu et al., 2015). However, the expensive cost and long detection time severely limit its mass adoption. Radiomics is a high-throughput analysis of quantitative tumor characteristics from standard-of-care medical imaging, like computed tomography (CT) and magnetic resonance imaging (MRI). By further modeling with machine learning, radiomics can provide better clinical-decision support systems for the clinicians, like tumor diagnosis and prognosis prediction (Lambin et al., 2017). Compared with genetic detection, radiographic testing is non-invasive and does little harm to the weak patients. Especially, comparing with MRI, CT is much cheaper, and its examination results can be available faster. As a preoperative routine test for CRC patients to locate the tumor before resection surgery, CT imaging analysis can provide timely guidance on surgical procedures and postoperative treatment. With sophisticated image processing tools to obtain high-dimensional image features, CT images contain abundant information which provides a powerful application in multiple medical studies (Limkin et al., 2017).

Typical radiomic features are mainly morphological characteristics of the tumor lesion, such as tumor size, shape and texture, which are customized according to human recognition cognition or compliant with certain human-defined rules (Gillies et al., 2016). The standardization of radiomic features makes it possible to quantify phenotypic characteristics on medical imaging (van Griethuysen et al., 2017). Through successfully applications in tumor differentiated grading (Kim et al., 2015), genomics prediction (Yang et al., 2018), prognosis predicting (Huang et al., 2016), evaluation of tumor immune microenvironment (Jiang et al., 2020) and prediction of chemoradiation therapy response (Shi et al., 2019), radiomic studies demonstrate that radiographic images can provide abundant information for cancer research. However, these radiomics features obtained by typical method are still limited by the human definition. It fails to consider the feature-to-feature relationship which plays a vital role in tumor microenvironment. Deep learning, one kind of machine learning based on artificial neural networks, has a powerful ability in image analysis (LeCun et al., 2015) with convolutional neural networks (CNNs). A few studies based on deep learning have proved its effectiveness in tumor assessment like lymph node status prediction (Zheng et al., 2020) and tumor recurrence prediction (Liu et al., 2022). Though with high prediction accuracy, deep learning is known as a black box as lacks the interpretation for its prediction, which makes it hard to be accepted by doctors. Excavating the hidden biological mechanism for the deep learning models will improve its interpretability and promote the clinical utility. Thus, there is an urgent need for developing a biologically interpretable deep learning model for predicting the prognosis of colorectal cancer.

In this study, we investigated an end-to-end CNNs model to quantify radiographic tumor characteristics and prognosis prediction for CRC patients. Deep features of CT images were extracted for correlation analysis with RNS-seq data from the ICGC-ARGO project (The International Cancer Genome Consortium-Accelerating Research in Genomic Oncology) to further explore the underlying biological mechanism learned by the Multi-Size Convolutional Neural Network (MSCNN) model. Our results proved that deep CT features can reveal the molecular information of tumors to some extent and ultimately improve the stratification of CRC patients.

MATERIALS AND METHODS

Patients and Data Collection

In this retrospective study, a total of 808 colorectal cancer patients who had cancer resection at the Sixth Affiliated Hospital of Sun Yat-sen University from 22 Jan 2008 to 30 Jan 2018 were included for analysis. Patients admitted during 2008-2013 were assigned to the development cohort ($n = 426$) for model construction and the rest of patients admitted during 2014-2018 were assigned to the validation cohort ($n = 382$) for model validation. All patients had CT examinations before the cancer resection surgery and the image data were stored in DICOM (Digital Imaging and Communications in Medicine) format. Region of interest (ROI) for colorectal cancer tumor area was manually delineated by experienced doctors with ITK-snap (Version 3.2) software. Baseline clinicopathological information containing age, gender, differentiated grade, lymph node metastasis and microsatellite status. Among these patients, 236 patients were enrolled in the ICGC-ARGO project and had paired RNA sequencing data.

Data Preprocessing and Enhancement

Figure 1 shows the pipeline of our analysis from origin CT images and their corresponding ROI to predict the disease-free survival for each patient. Origin CT images size is 512 * 512 with slices from 23 to 682 (mean = 162), and the valid slices which have tumor lesion of ROI for each CT image range between 3 and 77. To fit the deep learning model and reduce the computational

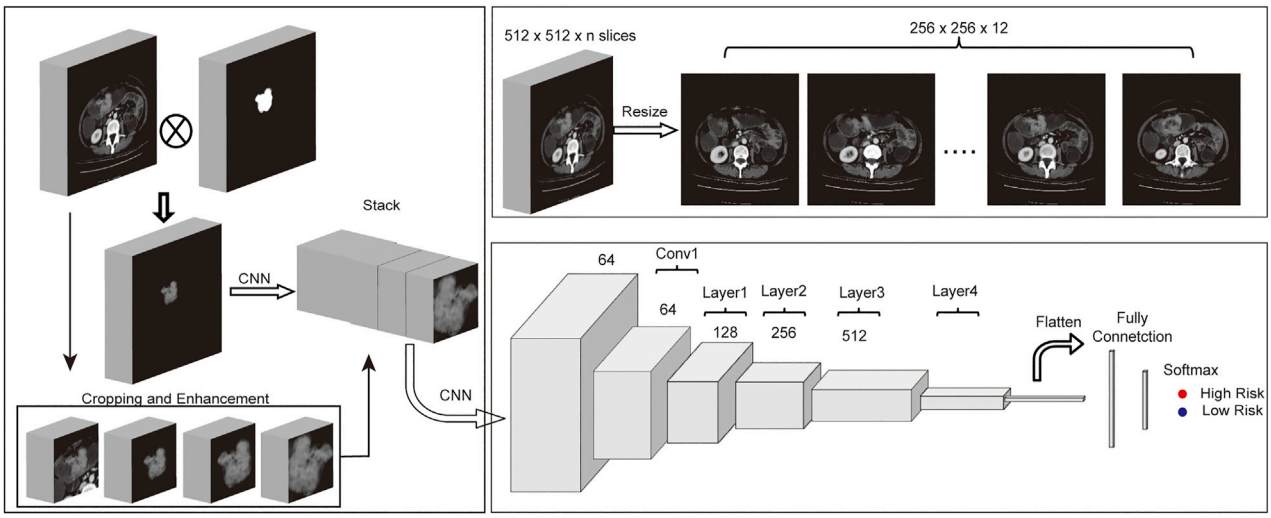


FIGURE 1 | Workflow of MSCNN. **(A)** Multi-Size based data enhancement of CT images before fed into MSCNN. **(B)** Data preprocessing of CT images with ROIs. **(C)** Network structure of MSCNN Multi-Size which includes a CNN to combine Multi-Size CT data, a ResNet34 network to extract image features of tumors from CT images and a last classification network.

parameters, all 3D CT images and ROIs only kept the slices with valid areas and then were resized to 256 * 256 * 12 with SciPy ndimage python submodule. To better conclude the tumor boundary information, all ROIs were binarily dilated with five pixels using morphology function in ndimage submodule. As the tumor ROI area of colorectal cancer is usually quite small, accounting for only 1–5% of the whole CT image, detailed information for the tumor is hard to extract from the deep learning model. To address this issue, the tumor area is cropped and magnified at different magnifications. Meanwhile, the cropped CT images were also augmented by rotating at random angles and flipping with a certain probability. Finally, all images for each patient were stacked together to feed into the neural network.

Multi-Size Convolutional Neural Network Model Construction

Convolutional Neural Network (CNN) is a powerful Deep Learning algorithm that can extract relevant texture features from the image. By stacking several CNNs, deep learning model can learn deeper features from the image according to the training task. Although model becomes much difficult to train if there are too many layers in deep learning networks, a residual neural network (ResNet (He et al., 2016)) is designed to solve this problem. Our model was based on ResNet34, which contains 34 convolutional neural networks and four residual blocks. First, one subnetwork with CNNs of different input sizes were designed for features extraction from the origin CT image and its enhanced cropped images. Then all features from these CNN were stacked together and following one CNN layer and the rest residual blocks of RenNet34 were used to extract higher and deeper features. Finally, one Fully Connected (FC) layer which contains one hidden layer with 64 nodes and one output layer finished the patient disease-free survival classification task.

Model Development and Validation

As shown in **Supplementary Figure S1**, for better training the model, only patients with tumor recurrence in 3 years or disease-free survival for more than 5 years were considered in the model development stage. CT images with ROI were fed into the deep learning model, and disease-free survival status was used as the labels. Model training was performed by updating the network weights using the backpropagation algorithm according to the cross-entropy loss between the prediction and the real outcomes. Adam optimizer was used in model network weights updating, and the learning rate was decayed to half for every 10 training epochs with an initial rate of 0.001. During training, the loss was continuously monitored, and model weights were saved when loss decreased. If the loss was not decreased for more than 20 epochs, then training was ended and saved model with the highest Area under the receiver operating characteristic (ROC) Curve (AUC) was loaded for further validation. CT signatures score was calculated on the whole development and validation cohort through the MSCNN model with CT images. A nomogram was constructed by incorporating the CT signature with other clinicopathologic risk factors, and its benefit was evaluated by the calibration curve and Decision curves analysis (DCA).

Radiomics Method

To compare our deep learning based method with conventional radiomics method, we constructed a model with CT radiomics features. For each of CT image, a total of 107 radiomics features were extracted using Pyradiomics (van Griethuysen et al., 2017) package in python 3.8 platform.

TABLE 1 | Baseline characteristic of patients in the development and validation cohort.

	level	Development cohort(*n* = 426)			Validation cohort (*n* = 328)		
		Low Risk	High Risk	P	Low Risk	High Risk	P
n		268	158		200	182	
Age (mean (SD))		58.732 (12.676)	59.816 (15.649)	0.4878	56.799 (13.099)	57.134 (13.191)	0.833
Sex (%)	F	116 (43.28)	57 (36.08)	0.1735	92 (46.00)	68 (37.36)	0.1085
	M	152 (56.72)	101 (63.92)		108 (54.00)	114 (62.64)	
TNM stage (%)	I	28 (10.45)	7 (4.43)	<0.0001	38 (19.19)	35 (19.34)	0.0026
	II	126 (47.01)	29 (18.35)		69 (34.85)	42 (23.20)	
	III	99 (36.94)	53 (33.54)		63 (31.82)	52 (28.73)	
	IV	15 (5.60)	69 (43.67)		28 (14.14)	52 (28.73)	
T stage (%)	T1	14 (5.22)	5 (3.18)	0.0001	9 (4.55)	8 (4.42)	0.6512
	T2	23 (8.58)	4 (2.55)		35 (17.68)	29 (16.02)	
	T3	208 (77.61)	112 (71.34)		133 (67.17)	117 (64.64)	
	T4	23 (8.58)	36 (22.93)		21 (10.61)	27 (14.92)	
N stage (%)	N0	157 (58.80)	51 (32.90)	<0.0001	118 (59.00)	84 (46.15)	0.0175
	N1	84 (31.46)	65 (41.94)		58 (29.00)	60 (32.97)	
	N2	26 (9.74)	39 (25.16)		24 (12.00)	38 (20.88)	
M stage (%)	M0	253 (94.40)	89 (56.33)	<0.0001	172 (88.21)	130 (78.31)	0.0168
	M1	15 (5.60)	69 (43.67)		23 (11.79)	36 (21.69)	
Differentiation grade (%)	Low	57 (30.81)	29 (26.36)	0.0256	45 (36.00)	29 (25.44)	0.181
	Moderate	117 (63.24)	64 (58.18)		76 (60.80)	79 (69.30)	
	High	11 (5.95)	17 (15.45)		4 (3.20)	6 (5.26)	
Chemotherapy Adjuvant (%)	No	68 (32.69)	32 (32.65)	1	108 (56.84)	102 (61.82)	0.3992
	Yes	140 (67.31)	66 (67.35)		82 (43.16)	63 (38.18)	

Standard Deviation (SD) and Median Absolute Deviation (MAD) were used to initially screen features with significant differences. Z-score normalization was performed to increase the comparability between the left radiomics features. The least absolute shrinkage and selection operator (LASSO) with cox regression was used to construct the final radiomics based model.

Deep Features Visualization

To visualize how the MSCNN divides patients into high recurrence risk and low recurrence risk, deep features from the last two layers of the MSCNN model were exported for further analysis. A correlation heatmap was performed on the 64 features from the hidden notes of the FC layer to show the most related deep CT features with high recurrence risk and low recurrence risk. Principal component analysis (PCA) analysis was performed on 512 origin deep CT features from the ResNet34 network and 64 features from hidden nodes of the FC layer.

Correlating the Computed Tomography Signature and Deep Computed Tomography Features With Gene Expression Data

To explore the biological characteristics of CT signature, Gene Ontology analysis and Gene Set Enrichment Analysis (GSEA) was conducted for differentially expressed genes between the risk groups. To further figure out how the model captures the underlying biological information from CT images, correlation analysis was performed between 64 deep CT features and cancer-related pathways. Functional spectra were calculated with the DeepCC method to explore the most related biological pathways with deep CT features (Gao et al., 2019). All hallmark pathways which have significant correlations with these 64 deep CT features were displayed in a bar plot.

Statistical Analyses

All statistical analyses were performed by R software (version 4.1.1). Kaplan-Meier curve was used to perform survival analysis for model prediction results with R package "survival". Log-rank test was used to evaluate results of the univariable analysis of model prediction results and other clinic-pathological factors with disease-free survival (DFS). Multivariable analysis was performed using the Cox proportional hazards regression method with only the significant variables from univariable analysis. Correlation analysis were performed using the Pearson method. For all analyses, the two-sided value p value < 0.05 was considered statistically significant.

RESULTS

Risk Prediction From Computed Tomography Images

We calculated the recurrence risk of colorectal cancer patients with CT images and ROI in an end-to-end deep learning method.

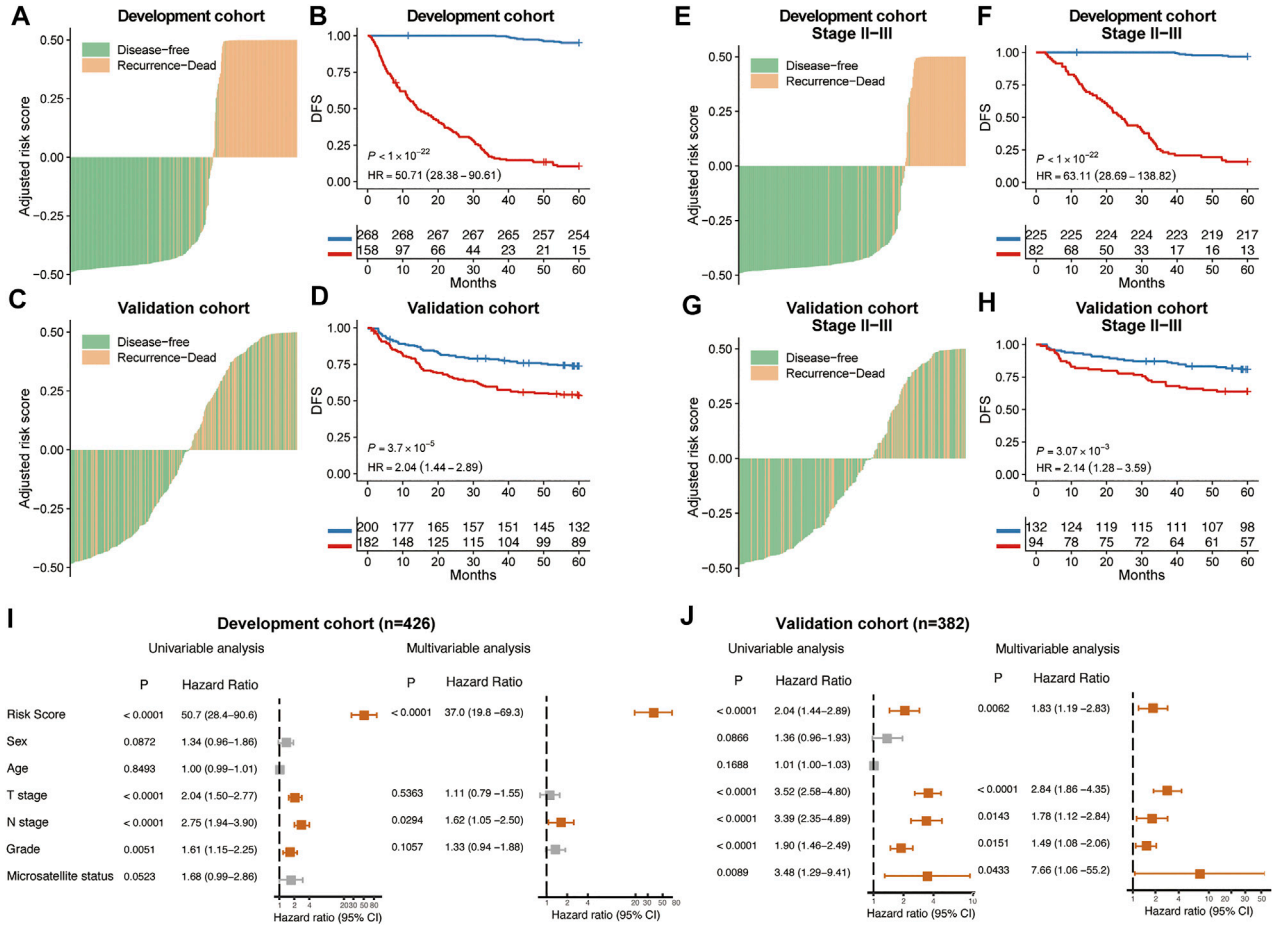


FIGURE 2 | Prognostic performance of MSCNN. The distribution of CT signature of MSCNN and its corresponding recurrence status in the development cohort **(A)** and validation cohort **(C)**. Kaplan-Meier curves showed a significant survival difference between the high and low risk groups in the development cohort **(B)** and validation cohort **(D)**. Prognostic analysis of CRC patients in stage II and III subgroups **(E–H)**. Univariable and multivariable analysis of clinical factors in the development cohort **(I)** and validation cohort **(J)**.

After model training with the development cohort, a CT signature score of each patient was calculated with the MSCNN model in **Supplementary Table S1**. Patients with a recurrence risk of more than 0.5 were classified into high risk groups, and the remain patients were in low risk group. Patients' clinical characteristics in development and validation cohort were displayed in **Table 1**.

High Risk and Low Risk Patients Show Significant Different Survival

In both development and validation cohorts, high risk patients show worse mean survival (23 vs. 105 months and 46 vs. 58 months). Kaplan-Meier curve revealed a significant association between CT images risk prediction and patients' DFS in the development cohort (HR: 50.7, 95% CI: 28.4–90.6, $p < 0.001$) and validation cohort (HR: 2.04, 95% CI: 1.44–2.89, $p < 0.001$) (**Figures 2A–D**). Previous research showed that clinicopathological information may be not enough to accurately predict the recurrence risk for colorectal patients with stage II and III (Tsikitis et al., 2014). Kaplan-Meier survival curve in stage II and III patients showed that risk prediction of our model can still divide those patients into significant survival different groups (**Figures 2E–H**). Univariable and multivariable cox regression analyses were performed to identity significant clinicopathological factors associated with cancer recurrence. Besides the risk scores calculated from CT images, clinical factors sex, age, T stage, N stage, differentiation grade and Microsatellite status were added to multivariable analysis. Forest plot showed risk scores from CT images was an independent prognostic predictor of cancer recurrence in both development and validation cohorts (**Figures 2I,J**).

Radiomics Model and Risk Prediction

Standard Deviation and Median Absolute Deviation for each radiomics features was calculated after z-score normalization and only 50 features with SD > 1 and MAD > 3.5 were left for subsequent modeling. Finally, 11 radiomics features were kept with LASSO-cox regression to construct the classification model. Radiomics score was calculated by a linear combination of non-zero coefficients multiplied with these 11 radiomics features. To classify high and low risk groups, the

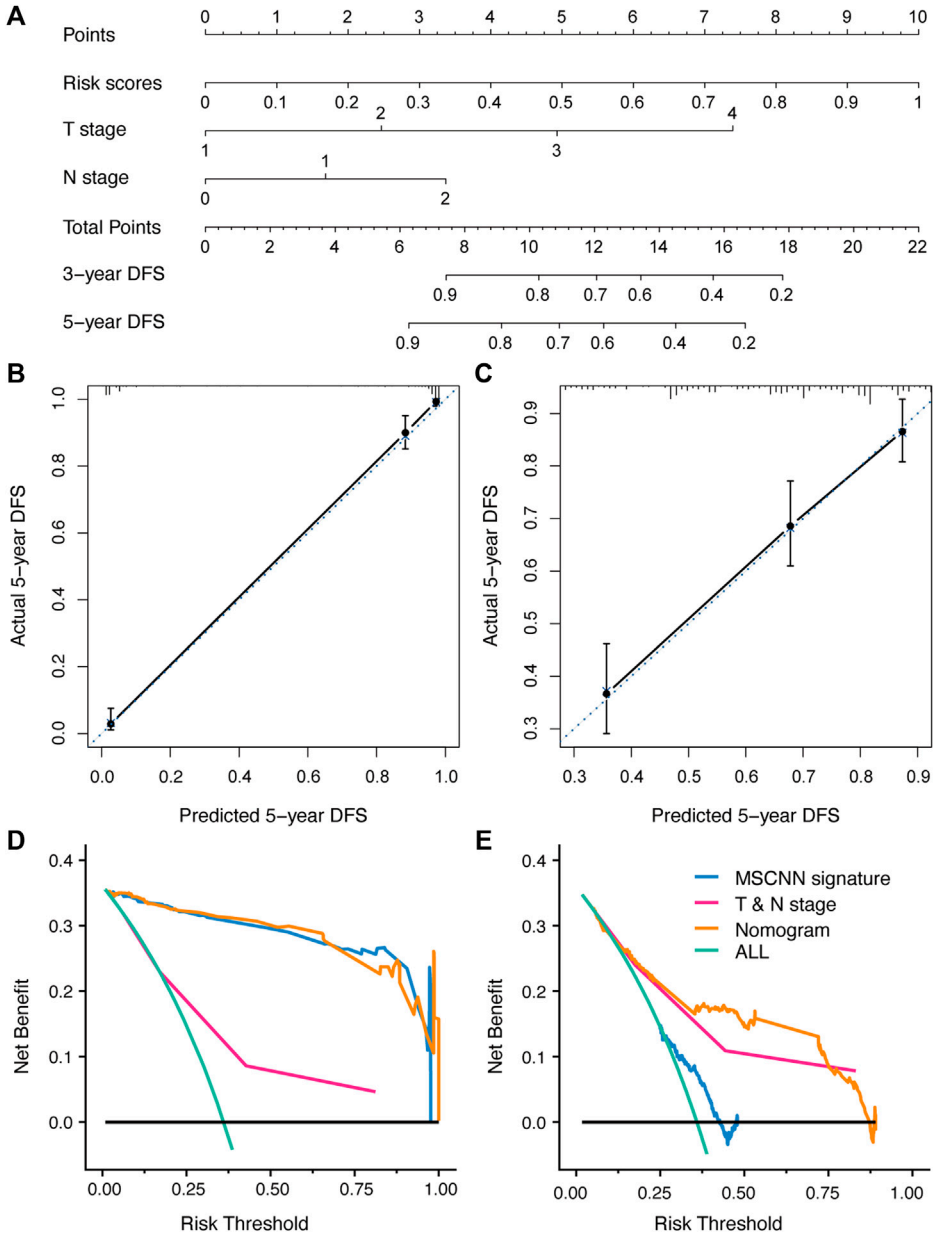


FIGURE 3 | The developed nomogram incorporated CT signature with T & N stage **(A)**. Coordinates length for each prognostic factor was determined by the coefficients of the cox regression model. For each patient, the total score was calculated with all variable scores. The probability of DFS was derived from the mapping relationship between the evaluation results and total score on specified patient survival time. **(B,C)** Calibration curves of nomogram for 5 years DFS in the development and validation cohort. **(D,E)** Decision curve analysis for nomogram established in the development and validation cohort.

optimal cut-off of radiomics scores was determined by the time-dependent ROC curve. Survival analysis showed significant differences between high risk patients and low risk patients according to radiomics scores (**Supplementary Figures S2B,D**). Comparison between our MSCNN method and Radiomics method were displayed with ROC curves and the result proved that our model could obtain better prediction of prognosis in both development and validation cohorts (**Supplementary Figures S2A,C**).

Nomogram for Risk Prediction From Radiomics

According to the multivariable analysis, the Cox regression model which incorporated CT signature, T stage and N stage was developed and displayed as a CT signature based nomogram (**Figure 3A**). The calibration curve of the radiomics nomogram showed good concordance between the prediction and the actual DFS survival (**Figures 3B,C**). DCA curve showed that nomogram achieved better net benefit compared with TNM-stage only (**Figures 3E,F**).

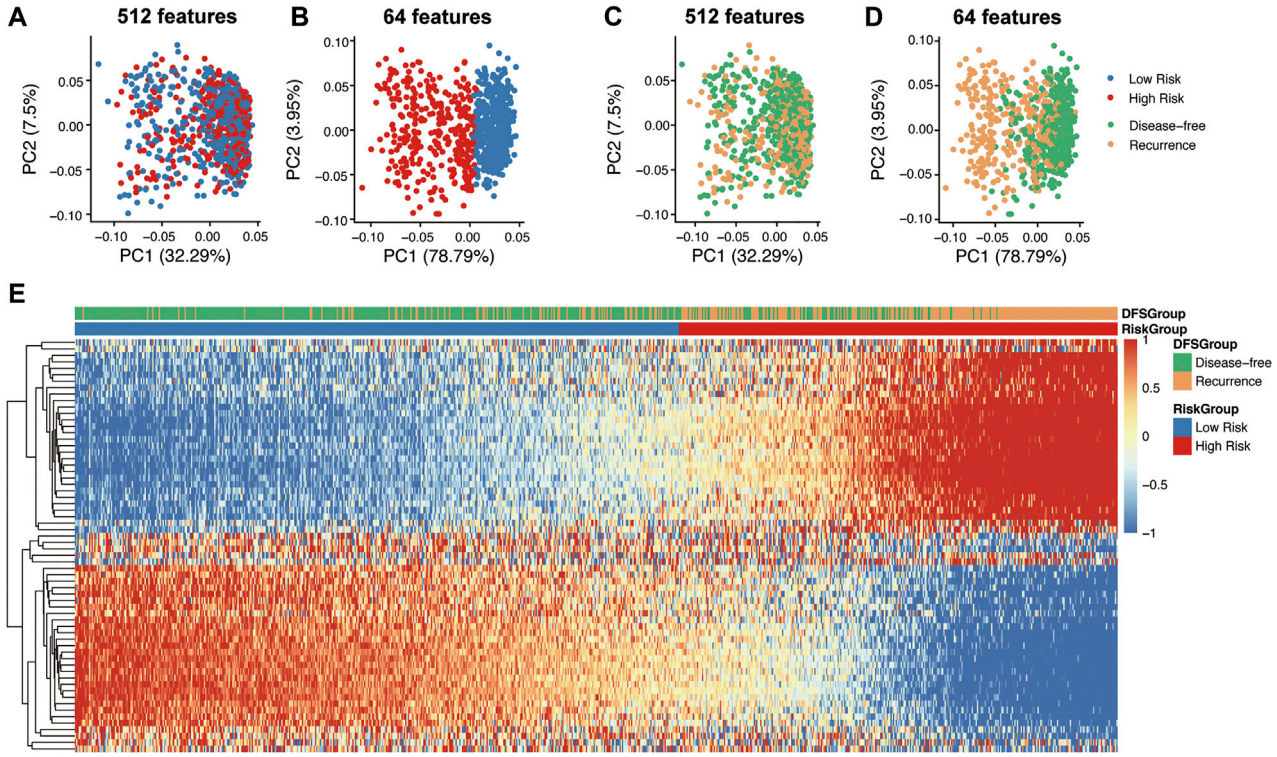


FIGURE 4 | Dimension reduction for visualization and correlation analysis of deep CT features. Principle component analysis (PCA) on the 512 features of the ResNet34 network **(A,C)** and 64 features (CT feature) of hidden notes of the FC network **(B,D)**. Correlation heatmap between 64 deep CT features and prognostic difference group **(E)**.

Visualization for the Deep Features From Radiomics

Deep features were extracted from the output of the ResNet34 network and hidden notes of the FC layer. 512 features were exported from the ResNet34 network for each CT image, and then the 64 features most related to tumor recurrence and disease-free survival were extracted from the hidden layer. PCA analysis showed deep features from the RseNet34 network were not enough to accurately divide patients into high risk groups and low risk groups (**Figures 4A,B**). However, recurrence related 64 features extracted from the hidden layers achieved distinct classification (**Figures 4C,D**). Unsupervised clustering of 64 deep CT features displayed in the heatmap showed that these deep CT features were significantly highly correlated with high and low risk subgroups of CRC patients (**Figure 4E**).

Pathway Analysis of Radiomics Risk Group and Deep Features

To further explore the biological interpretability of deep CT features from the MSCNN model, Gene Ontology analysis of the different groups and the GSEA showed significant enrichment of immune pathways (**Figure 5A**) such as Interferon alpha response ($p < 0.001$), Interferon Gamma Response ($p < 0.001$) and Inflammatory response ($p = 0.037$) (**Figure 5B**). Significantly differential expression genes of risk groups were shown in **Supplementary Figure S3**. Besides, correlation analysis of the 64 deep CT features (**Figure 5C**) found that most of these features were highly correlated. Their further correlation analysis with the hallmark pathways was performed to explore the biological mechanism of the MSCNN model. Hallmark pathways were selected according to significant association with those deep CT features, and the result showed those features had a significant enrichment in some metabolism and proliferative pathways (**Figure 5D**).

DISCUSSION

In this study, we proposed a deep learning based end-to-end method to predict prognosis of colorectal cancer patients after tumor resection surgery from CT images. Our deep learning model successfully screened out high tumor recurrence risk patients with significant prognostic differences from the others. Univariable and multivariable analyses showed that CT signature was an independent factor for CRC patient survival prediction. By incorporating CT signature and clinical risk factors, we built a nomogram that can facilitate the risk prediction for colorectal cancer patients. Correlation analysis with genomic data indicated that high risk patients showed downregulation of immune pathways and deep CT features learned by MSCNN model were significantly enriched in some metabolism and proliferative pathways.

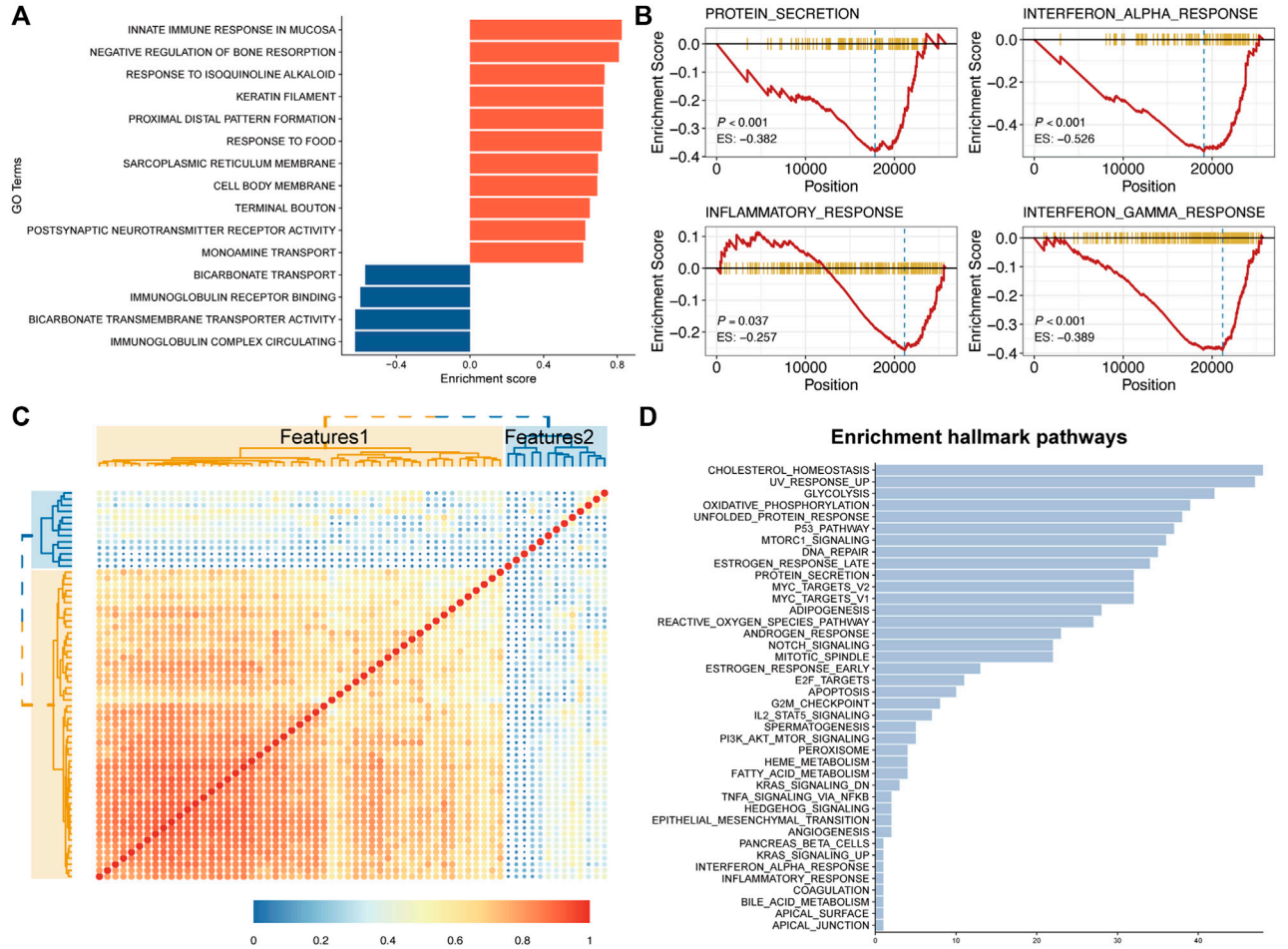


FIGURE 5 | Global gene set pathway analysis. **(A)** Gene Ontology pathway enrichment analysis between CT signatures and RNA-Seq expression. **(B)** GSEA showed several Immune related pathways were downregulated in high CT signature patients. **(C,D)** Correlation between 64 deep CT features and their enrichment hallmark pathways.

Traditional prognostic analysis based on genetic testing can obtain good performance as expression of several genes were highly related with patient's tumor progression (Kandimalla et al., 2018; Sveen et al., 2020). However high cost and long test time cycle limited its large-scale applications. Compared with genetic testing, CT imaging, a much cheaper non-invasive preoperative routine test for CRC patients to locate the tumor before the resection surgery, can provide more preoperative interventions. Our study was based on deep learning model which focused detailed and deeper information of CT images and acquired good performance in prognostic prediction for CRC patients.

Deep learning model with CNN can learn the features of CT images from low to high dimensions and their correlation (Yamashita et al., 2018), which may be the key reason for high performance in image analysis. Since most previous CT image based prognostic research have only used pretrained deep learning to extract images features, subsequent analysis required subjective screening of these features to build the machine learning model again (Huang et al., 2020; Park et al., 2021; Liu et al., 2022). Besides, they did not consider the special characteristics of medical images which mean generic pretrained deep learning models were not suitable. Our MSCNN model was an end-to-end method to quickly predict the prognosis for CRC patients with CT images, which can also reduce the subjectivity of human selection of image characteristics. In addition, the percentage of tumors in CT images is often small, accounting for only about 1–5%, which makes it difficult for ordinary CNN models to learn the key information of CT images. Based on the idea of multi-instance learning in pathology research (Bilal et al., 2021; Sirinukunwattana et al., 2021), our MSCNN model considered both full-image and local detail information of CT images by cropping and deflating the ROI region, making the prognostic predicting of our model more comprehensive and accurate.

Recent rapid development of deep learning has generated a series of CNN based studies for radiographic analysis, like treatment response predicting (Xu et al., 2019; Lu et al., 2021) and detection of Synchronous Peritoneal Carcinomatosis (Yuan et al., 2020). However, few of them considered the interpretation of their deep learning models, making it hard for clinicians to be convinced

of their findings. Our study not only visualized the process of classifying CRC patients in high and low risk groups, but also found that the CT signature of our MSCNN model was significantly correlated with several immune pathways. Meanwhile, our results found that deep CT features showed significant enrichment in some metabolic proliferative pathways which was consistent with previous studies (Kandimalla et al., 2019; Cai et al., 2020; La Vecchia and Sebastián, 2020).

However, despite satisfactory results with sufficient biological interpretation, our study still has some limitations. First, a prospective study was needed to further confirm and optimize our model. In our study, all patients included are from one single center, which may cause bias for the model validation. In addition, our CT images for prognosis predicting need manual ROI segmentation which is time-consuming and seriously affects the applicability of our model. This can be achieved by object detection through deep learning with enough data. In this way, the ROI can be directly learned from the model without manual sketching.

In conclusion, our study demonstrated that deep learning with CT images can be effectively applied to cancer recurrence prediction. By incorporating clinical factors, more accurate results can be achieved than just routine TNM staging. Correlation analysis with gene expression data showed that deep CT features captured by our model did have a biological meaning which gave credibility to our MSCNN model.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this study are available for reasonable request from the corresponding authors (FG). The CT images data are not public as they contain some private information of our patients. We only used gene expression data (but not microarray/deep sequencing) for pathway analysis to enhance the interpretability of the main results. Gene expression data of the ICGC-ARGO project are not public available currently and will be released on the website of the project (https://www.icgc-argo.org/page/114/cgcc) by ICGC after the milestone.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Medical Ethics Committee of the Sixth Affiliated Hospital of Sun Yat-sen University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

FG and XW designed the whole study. C-HL and DC wrote the paper. C-HL and DC analyzed and interpreted the data and drew the figures. DC, FG, and XW revised the paper. M-EZ planned all the data for this study. M-EZ, Z-PH, and QZ delineated the ROIs. M-YL, CH, and HQ collected and cleaned up the clinical data. COCC Working Group helped with genomic sequencing and analysis. All authors contributed to manuscript revision and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 82002221, FG and No. 82102475, M-EZ), Guangzhou Basic and Applied Basic Research Fund (No. 202102020820, FG), The Sixth Affiliated Hospital of Sun Yat-sen University Start-up Fund for Returnees (No. R20210217202501975, FG), China Postdoctoral Science Foundation (No. 2020M683121, M-EZ), the Guangdong Basic and Applied Basic Research Foundation (No. 2020A1515110489, M-EZ), the China Postdoctoral Science Foundation (No. 2021T140769, M-EZ).

ACKNOWLEDGMENTS

Thanks to the COCC Working Group for their help with this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.880093/full#supplementary-material

REFERENCES

Bilal, M., Raza, S. E. A., Azam, A., Graham, S., Ilyas, M., Cree, I. A., et al. (2021). Development and Validation of a Weakly Supervised Deep Learning Framework to Predict the Status of Molecular Pathways and Key Mutations in Colorectal Cancer from Routine Histology Images: a Retrospective Study. *Lancet Digital Health* 3 (12), e763–e772. doi:10.1016/s2589-7500(21)00180-1

Cai, D., Duan, X., Wang, W., Huang, Z. P., Zhu, Q., Zhong, M. E., et al. (2020). A Metabolism-Related Radiomics Signature for Predicting the Prognosis of Colorectal Cancer. *Front. Mol. Biosci.* 7, 613918. doi:10.3389/fmolb.2020.613918

Gao, F., Wang, W., Tan, M., Zhu, L., Zhang, Y., Fessler, E., et al. (2019). DeepCC: a Novel Deep Learning-Based Framework for Cancer Molecular Subtype Classification. *Oncogenesis* 8 (9), 44. doi:10.1038/s41389-019-0157-8

Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* 278 (2), 563–577. doi:10.1148/radiol.2015151169

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016, 770–778.

Huang, S.-C., Kothari, T., Banerjee, I., Chute, C., Ball, R. L., Borus, N., et al. (2020). PENet-A Scalable Deep-Learning Model for Automated Diagnosis of

Pulmonary Embolism Using Volumetric CT Imaging. *npj Digit. Med.* 3, 61. doi:10.1038/s41746-020-0266-y

Huang, Y., Liu, Z., He, L., Chen, X., Pan, D., Ma, Z., et al. (2016). Radiomics Signature: A Potential Biomarker for the Prediction of Disease-free Survival in Early-Stage (I or II) Non-small Cell Lung Cancer. *Radiology* 281 (3), 947–957. doi:10.1148/radiol.2016152234

Jiang, Y., Wang, H., Wu, J., Chen, C., Yuan, Q., Huang, W., et al. (2020). Noninvasive Imaging Evaluation of Tumor Immune Microenvironment to Predict Outcomes in Gastric Cancer. *Ann. Oncol.* 31 (6), 760–768. doi:10.1016/j.annonc.2020.03.295

Joachim, C., Macni, J., Drame, M., Pomier, A., Escarmant, P., Veronique-Baudin, J., et al. (2019). Overall Survival of Colorectal Cancer by Stage at Diagnosis. *Med. Baltim.* 98 (35), e16941. doi:10.1097/md.0000000000016941

Kandimalla, R., Gao, F., Matsuyama, T., Ishikawa, T., Uetake, H., Takahashi, N., et al. (2018). Genome-wide Discovery and Identification of a Novel miRNA Signature for Recurrence Prediction in Stage II and III Colorectal Cancer. *Clin. Cancer Res.* 24 (16), 3867–3877. doi:10.1158/1078-0432.ccr-17-3236

Kandimalla, R., Ozawa, T., Gao, F., Wang, X., Goel, A., Nozawa, H., et al. (2019). Gene Expression Signature in Surgical Tissues and Endoscopic Biopsies Identifies High-Risk T1 Colorectal Cancers. *Gastroenterology* 156 (8), 2338–2341. doi:10.1053/j.gastro.2019.02.027

Kim, J. E., Lee, J. M., Baek, J. H., Moon, S. K., Kim, S. H., Han, J. K., et al. (2015). Differentiation of Poorly Differentiated Colorectal Adenocarcinomas from Well- or Moderately Differentiated Colorectal Adenocarcinomas at Contrast-Enhanced Multidetector CT. *Abdom. Imaging* 40 (1), 1–10. doi:10.1007/s00261-014-0176-z

La Vecchia, S., and Sebastián, C. (2020). Metabolic Pathways Regulating Colorectal Cancer Initiation and Progression. *Seminars Cell & Dev. Biol.* 98, 63–70. doi:10.1016/j.semcdb.2019.05.018

Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., et al. (2017). Radiomics: the Bridge between Medical Imaging and Personalized Medicine. *Nat. Rev. Clin. Oncol.* 14 (12), 749–762. doi:10.1038/nrclinonc.2017.141

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Limkin, E. J., Sun, R., Dercle, L., Zacharaki, E. I., Robert, C., Reuzé, S., et al. (2017). Promises and Challenges for the Implementation of Computational Medical Imaging (Radiomics) in Oncology. *Ann. Oncol.* 28 (6), 1191–1206. doi:10.1093/annonc/mdx034

Liu, S., Sun, W., Yang, S., Duan, L., Huang, C., Xu, J., et al. (2022). Deep Learning Radiomic Nomogram to Predict Recurrence in Soft Tissue Sarcoma: a Multi-Institutional Study. *Eur. Radiol.* 32 (2), 793–805. doi:10.1007/s00330-021-08221-0

Lu, L., Dercle, L., Zhao, B., and Schwartz, L. H. (2021). Deep Learning for the Prediction of Early On-Treatment Response in Metastatic Colorectal Cancer from Serial Medical Imaging. *Nat. Commun.* 12 (1), 6654. doi:10.1038/s41467-021-26990-6

Molinari, C., Marisi, G., Passardi, A., Matteucci, L., De Maio, G., and Ulivi, P. (2018). Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? *Int. J. Mol. Sci.* 19 (12), 3733. doi:10.3390/ijms19123733

Nagtegaal, I. D., Quirke, P., and Schmoll, H.-J. (2011). Has the New TNM Classification for Colorectal Cancer Improved Care? *Nat. Rev. Clin. Oncol.* 9 (2), 119–123. doi:10.1038/nrclinonc.2011.157

Park, Y. J., Choi, D., Choi, J. Y., and Hyun, S. H. (2021). Performance Evaluation of a Deep Learning System for Differential Diagnosis of Lung Cancer with Conventional CT and FDG PET/CT Using Transfer Learning and Metadata. *Clin. Nucl. Med.* 46 (8), 635–640. doi:10.1097/RLU.0000000000003661

Pierotti, M. A. (2017). The Molecular Understanding of Cancer: from the Unspeakable Illness to a Curable Disease. *ecancer* 11, 747. doi:10.3332/ecancer.2017.747

Shi, L., Zhang, Y., Nie, K., Sun, X., Niu, T., Yue, N., et al. (2019). Machine Learning for Prediction of Chemoradiation Therapy Response in Rectal Cancer Using Pre-treatment and Mid-radiation Multi-Parametric MRI. *Magn. Reson. Imaging* 61, 33–40. doi:10.1016/j.mri.2019.05.003

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal Cancer Statistics, 2020. *CA A Cancer J. Clin.* 70 (3), 145–164. doi:10.3322/caac.21601

Sirinukunwattana, K., Domingo, E., Richman, S. D., Redmond, K. L., Blake, A., Verrill, C., et al. (2021). Image-based Consensus Molecular Subtype (imCMS) Classification of Colorectal Cancer Using Deep Learning. *Gut* 70 (3), 544–554. doi:10.1136/gutjnl-2019-319866

Sorbye, H., Köhne, C.-H., Sargent, D. J., and Glimelius, B. (2007). Patient Characteristics and Stratification in Medical Treatment Studies for Metastatic Colorectal Cancer: a Proposal for Standardization of Patient Characteristic Reporting and Stratification. *Ann. Oncol.* 18 (10), 1666–1672. doi:10.1093/annonc/mdm267

Sveen, A., Kopetz, S., and Lothe, R. A. (2020). Biomarker-guided Therapy for Colorectal Cancer: Strength in Complexity. *Nat. Rev. Clin. Oncol.* 17 (1), 11–32. doi:10.1038/s41571-019-0241-1

Tsikitis, V. L., Larson, D. W., Huebner, M., Lohse, C. M., and Thompson, P. A. (2014). Predictors of Recurrence Free Survival for Patients with Stage II and III Colon Cancer. *BMC Cancer* 14, 336. doi:10.1186/1471-2407-14-336

van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 77 (21), e104–e107. doi:10.1158/0008-5472.can-17-0339

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., and Kerr, D. (2009). Genetic Prognostic and Predictive Markers in Colorectal Cancer. *Nat. Rev. Cancer* 9 (7), 489–499. doi:10.1038/nrc2645

Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., et al. (2019). Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clin. Cancer Res.* 25 (11), 3266–3275. doi:10.1158/1078-0432.ccr-18-2495

Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional Neural Networks: an Overview and Application in Radiology. *Insights Imaging* 9 (4), 611–629. doi:10.1007/s13244-018-0639-9

Yang, L., Dong, D., Fang, M., Zhu, Y., Zang, Y., Liu, Z., et al. (2018). Can CT-based Radiomics Signature Predict KRAS/NRAS/BRAF Mutations in Colorectal Cancer? *Eur. Radiol.* 28 (5), 2058–2067. doi:10.1007/s00330-017-5146-8

Yu, J., Wu, W. K. K., Li, X., He, J., Li, X.-X., Ng, S. S. M., et al. (2015). Novel Recurrently Mutated Genes and a Prognostic Mutation Signature in Colorectal Cancer. *Gut* 64 (4), 636–645. doi:10.1136/gutjnl-2013-306620

Yuan, Z., Xu, T., Cai, J., Zhao, Y., Cao, W., Fichera, A., et al. (2020). Development and Validation of an Image-Based Deep Learning Algorithm for Detection of Synchronous Peritoneal Carcinomatosis in Colorectal Cancer. *Ann. Surg.* 275 (4), e645–e651. doi:10.1097/sla.0000000000004229

Zheng, X., Yao, Z., Huang, Y., Yu, Y., Wang, Y., Liu, Y., et al. (2020). Deep Learning Radiomics Can Predict Axillary Lymph Node Status in Early-Stage Breast Cancer. *Nat. Commun.* 11 (1), 1236. doi:10.1038/s41467-020-15027-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, Cai, Zhong, Lv, Huang, Zhu, Hu, Qi, Wu and Gao. 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 Physiology

ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fphys.2022.909277

S-Thyroid Computer-Aided Diagnosis Ultrasound System of Thyroid Nodules: Correlation Between Transverse and Longitudinal Planes

Keen Yang[1†], Jing Chen[2†], Huaiyu Wu[2], Hongtian Tian[2], Xiuqin Ye[2], Jinfeng Xu[2], Xunpeng Luo[3]* and Fajin Dong[2]**

[1]The Second Clinical Medical College, Jinan University, Shenzhen, China, [2]Department of Ultrasound, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, China, [3]Department of Thyroid Surgery, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, China*

OPEN ACCESS

Edited by:
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

Reviewed by:
Man Lu,
Sichuan Cancer Hospital, China
Cui Li Gang,
Peking University Third Hospital, China

***Correspondence:**
Jinfeng Xu
xujinfeng@yahoo.com
Xunpeng Luo
luoxp4301@163.com
Fajin Dong
dongfajin@szhospital.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

Received: *31 March 2022*
Accepted: *13 April 2022*
Published: *20 May 2022*

Citation:
Yang K, Chen J, Wu H, Tian H, Ye X, Xu J, Luo X and Dong F (2022) S-Thyroid Computer-Aided Diagnosis Ultrasound System of Thyroid Nodules: Correlation Between Transverse and Longitudinal Planes. Front. Physiol. 13:909277. doi: 10.3389/fphys.2022.909277

Introduction: We compare the differences in the diagnostic results of S-thyroid, a computer-aided diagnosis (CAD) software, based on two mutually perpendicular planes.

Methods: Initially, 149 thyroid nodules confirmed by surgical pathology were enrolled in our study. CAD in our study was based on the ACR TI-RADS lexicon. *t* test, rank-sum test, and Chi-square test were used. The interclass correlation coefficient and Cohen's kappa were used to explore the correlation between CAD features. Receiver operating characteristic was plotted for different combinations of CAD features.

Results: The patient's age, transverse diameter, longitudinal diameter, shape, margin, echogenicity, echogenic foci, composition, TI-RADS classification, and risk probability of nodules in the transverse and longitudinal planes were related to thyroid cancer ($p < 0.05$). The AUC (95%CI) of TI-RADS classification in the transverse plane of CAD is better than that of the longitudinal plane [0.90 (0.84–0.95) vs. 0.83 (0.77–0.90), $p = 0.04$]. The AUC (95%CI) of risk probability of nodules in the transverse planes shows no difference from that in the longitudinal plane statistically [0.90 (0.85–0.95) vs. 0.88 (0.82–0.94), $p = 0.52$]. The AUC (95% CI), specificity, sensitivity, and accuracy [TI-RADS classification (transverse plane) + TI-RADS classification (longitudinal plane) + risk (transverse plane) + risk (longitudinal plane)] are 0.93 (0.89–0.97), 86.15%, 90.48%, and 88.59%, respectively.

Conclusion: The diagnosis of thyroid cancer in the CAD transverse plane was superior to that in the CAD longitudinal plane when using the TI-RADS classification, but there was no difference in the diagnosis between the two planes when using risk. However, the combination of CAD transverse and longitudinal planes had the best diagnostic ability.

Keywords: computer-aided diagnosis, ultrasound, thyroid nodules, S-thyroid, transverse, longitudinal

Abbreviations: US, ultrasound; CAD, computer-aided diagnosis; PD, plane difference; PD2, plane difference factor; ROC, receiver operating characteristic; AUC, area under ROC curve; TD, transverse diameter, LD, longitudinal diameter; AD, anteroposterior diameter; ICC, interclass correlation coefficient; TI-RADS, Thyroid Imaging Reporting and Data System; and ACR TI-RADS, American College of Radiology, Thyroid Imaging Reporting and Data System.

INTRODUCTION

Epidemiological studies show that thyroid cancer accounts for 3% of new cancers in women, with 32,130 cases compared to 12,150 cases in men. (Siegel et al., 2021). The increase in the number of thyroid cancer diagnoses is due in large part to the increasing use of diagnostic imaging technology and medical surveillance, as well as improved access to health care, all of which facilitate the detection of small, subclinical thyroid nodules and small thyroid cancers (Grani et al., 2020; Nambron et al., 2020).

Thyroid ultrasound (US) is the most effective tool for detecting thyroid lesions, especially when remnants of normal thyroid tissue are present, compared to other imaging studies such as computed tomography and magnetic resonance imaging (Grani and Fumarola, 2014; Hoang et al., 2018; Filetti et al., 2019). However, the repeatability and objectivity of the US are low, for the US highly dependent on operator experience and does not allow the analysis of image features quantitatively (Lee et al., 2016; Persichetti et al., 2020; Zhang et al., 2021).

To improve diagnostic accuracy, computer-aided diagnosis (CAD) systems have been developed (Shen et al., 2019; Zhang et al., 2021). CAD systems allow for quantitative assessment by efficiently analyzing large numbers of images, a computer-based approach that facilitates interpretation and diagnosis, and also reduces intra- and inter-observer variability (Singh et al., 2011). S-thyroid, similar to S-detect, is a computer-aided diagnostic software for ultrasound identification and differentiation of benign and malignant thyroid nodules. Some studies have investigated the diagnostic value of S-detect (Samsung Medison Co., Seoul, South Korea) for benign and malignant thyroid nodules (Choi et al., 2017; Kim et al., 2019; Xia et al., 2019; Barczyński et al., 2020; Wei et al., 2020), but no studies have been done on the diagnostic accuracy of S-thyroid for thyroid nodules. What is more, whether it is S-detect or S-thyroid, their diagnoses are based on a single ultrasound image. This is different from an ultrasonographer, who determines the benignity or malignancy of a thyroid nodule based on a combination of information from the transverse and longitudinal views of the thyroid nodule. However, no studies have yet examined the diagnostic variability of CAD between two mutually perpendicular views of the thyroid.

Consequently, the purpose of this study was to explore the differences between the diagnosis of thyroid nodules based on two mutually perpendicular planes of the S-thyroid software and the diagnostic efficacy of S-thyroid.

MATERIALS AND METHODS

Informed Consent

This retrospective study was approved by the institutional ethics committee of our hospital.

Patients

This retrospective study was approved by the appropriate institutional and research ethics committee. The inclusion and exclusion criteria are listed as follows:

1) US and CAD can detect the thyroid nodules of a patient.
2) Thyroid nodules range from 2 to 50 mm.
3) Thyroid nodule pathology was finally confirmed by surgical pathology.
4) Other non-thyroid cancers, such as lymphoma and metastatic cancers, were excluded.

The flow chart is shown in **Figure 1**.

US AND CAD DETECTION

The thyroid US detection was performed by an ultrasonographer with 15 years of experience in thyroid US detection using a 3–17 MHz linear array probe and a real-time US system (SonoScape Medical Corp., Shenzhen, Guangdong Province, China). First, the ultrasonographer performed a transverse scan to observe the entire thyroid gland, followed by a longitudinal scan. The transverse plane and longitudinal plane of the thyroid nodules with the most malignant signs have been preserved sequentially. Ultimately, the ultrasonographer measured the transverse diameter (TD), longitudinal diameter (LD), and anteroposterior diameter (AD) without knowing the pathology and CAD results.

CAD in our study was based on the American College of Radiology, Thyroid Imaging Reporting and Data System (ACR TI-RADS) lexicon (Tessler et al., 2018), and using S-thyroid software (SonoScape Medical Corp., Shenzhen, Guangdong Province, China). The thyroid nodules included in this study were not utilized for prior training or validation of the CAD system.

The CAD data in our study were obtained by the same ultrasonographer using the preserved images in both transverse and longitudinal planes. First, open S-thyroid, then without the need to outline the nodules manually, the CAD will automatically outline the

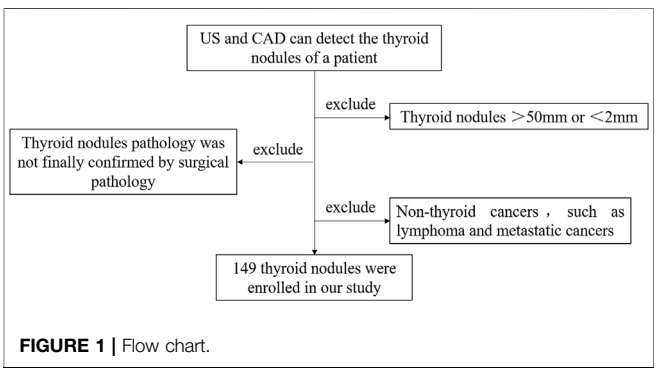


FIGURE 1 | Flow chart.

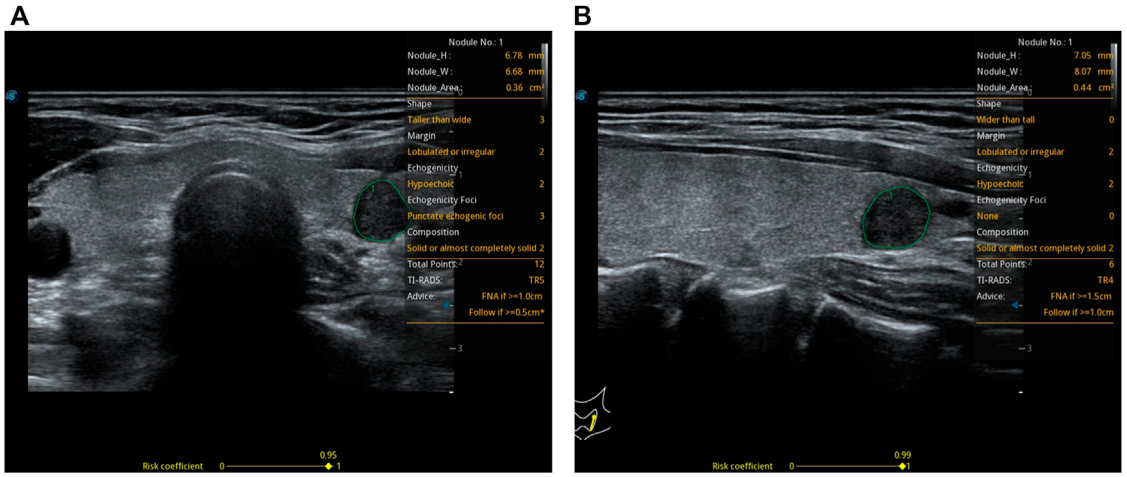


FIGURE 2 | A 37-year-old woman with a thyroid nodule and pathological findings suggestive of a micro papillary thyroid carcinoma. S-thyroid analyzes the features of the lesion. **(A)** Transverse plane: shape = taller-than-wide; margin = lobulated or irregular; echogenicity = hypoechoic; echogenic foci = punctate echogenic foci; composition = solid or almost completely solid; TI-RADS classification = TR5; risk = 0.95. **(B)** Longitudinal plane: shape = wider-than-tall; margin = lobulated or irregular; echogenicity = hypoechoic; echogenic foci = none; composition = solid or almost completely solid; TI-RADS classification = TR4; risk = 0.99.

thyroid nodule margin and display the following automatically as follows:

- TD, LD, AD;
- Composition: cystic or almost = 0; spongiform = 0; mixed cystic and solid = 1; solid or almost completely solid = 2;
- Echogenicity: anechoic = 0; hyperechoic or isoechoic = 1; hypoechoic = 2; very hypoechoic = 3;
- Shape: wider-than-tall = 0; taller-than-wide = 3;
- Margin: smooth = 0; ill-defined = 0; lobulated or irregular = 2; extra-thyroidal extension = 3;
- Echogenic Foci: none or large = 0; macrocalcifications = 1; peripheral (rim) calcifications = 2; punctate echogenic foci = 3;
- TI-RADS: TR1 = 0; TR2 = 2; TR3 = 3; TR4 = 4–6; TR5≥7.
- Risk: the CAD system assigns a score of 0–1, representing an increasing probability of malignancy in our study.

The relevant cases of CAD are shown in **Figure 2**.

What is more, two other sonographers with 5 and 10 years of experience in thyroid ultrasound detection, respectively, also read the saved images in the transverse and longitudinal planes and then graded the thyroid nodules according to ACR TI-RADS guidelines.

Difference Between Transverse and Longitudinal Planes

To evaluate the difference between the transverse plane (view T) and longitudinal plane (view L), we calculated the plane difference (PD) and plane difference factor (PD^2) through the equations given below:

$$PD_{composition} = composition_{view\ L} - composition_{view\ T}$$
$$PD_{echogenicity} = echogenicity_{view\ L} - echogenicity_{view\ T}$$
$$PD_{shape} = shape_{view\ L} - shape_{view\ T}$$
$$PD_{margin} = margin_{view\ L} - margin_{view\ T}$$
$$PD_{echogenic\ foci} = echogenic\ foci_{view\ L} echogenic\ foci_{view\ T}$$
$$PD_{risk} = risk\ score_{view\ L} - risk\ score_{view\ T}$$
$$PD^2 = PD * PD$$

Clinical Findings

Patients' gender, age, pathology results, and the location of the nodules were recorded. We divided the thyroid gland into three parts, including the left lobe, right lobe, and isthmus.

Statistical Analysis

The statistical analysis was performed by R (https://www.r-project.org) and IBM SPSS 25, and the figures were assembled with Adobe Illustrator CS6 and GraphPad Prism 8. *t* test was used for the normally distributed numerical variables, the rank-sum test was used for the non-normally distributed numerical variables, and the Chi-square test was used for the disordered classification variables. $p < 0.05$, as standard, statistically significant variables were included for further study.

The result consistency of numerical variables between the transverse and longitudinal planes of CAD was analyzed by the interclass correlation coefficient (ICC) while the result consistency of disordered classification variables was analyzed by Cohen's kappa. Multivariate logistic regression was used to construct the model, and then the ROC was plotted based on the results of multivariate logistic regression, to combine different CAD characteristics.

TABLE 1 | Patients' basic information.

Variable	Total (n = 149)	Benign (n = 65)	Malignant (n = 84)	p
Gender[c]				0.53
Male	25 (17)	9 (14)	16 (19)	
Female	124 (83)	56 (86)	68 (81)	
Age[d]	44 (36, 54	46 (41, 55	42 (34.75, 51.25	0.02
Location[c]				1
Left Lobe	69 (46)	30 (46)	39 (46)	
Right Lobe	78 (52)	34 (52)	44 (52)	
Isthmus	2 (1)	1 (2)	1 (1)	
Pathology				
Hashimoto thyroiditis		7 (11%)		
Nodular goiter		48 (74%)		
Follicular adenoma		8 (12%)		
Thyroid Hurthle cell adenoma		2 (3%)		
Papillary carcinoma			34 (40%)	
Micropapillary carcinoma			50 (60%)	
TI-RADS classification[a, d]				<0.01
1	19 (13)	17 (26)	2 (2)	
2	19 (13)	18 (28)	1 (1)	
3	10 (7)	8 (12)	2 (2)	
4	28 (19)	16 (25)	12 (14)	
5	73 (49)	6 (9)	67 (80)	
Risk[a, d]	0.75 (0.15, 0.95	0.14 (0.07, 0.37	0.94 (0.81, 0.97	<0.01
TI-RADS classification[b, c]				<0.01
1	17 (11)	16 (25)	1 (1)	
2	20 (13)	15 (23)	5 (6)	
3	11 (7)	8 (12)	3 (4)	
4	26 (17)	15 (23)	11 (13)	
5	75 (50)	11 (17)	64 (76)	
Risk[b, d]	0.87 (0.17, 0.97	0.21 (0.07, 0.57	0.96 (0.92, 0.98	<0.01

[a]*Based on the CAD of the transverse plane.*
[b]*Based on the CAD of the longitudinal plane.*
[c]*Non-normally distributed numerical variables are shown by median (first quartile, third quantile).*
[d]*Disordered classification variables are shown by percentage.*

RESULTS

Patients' Sample

There were a total of 149 thyroid nodules (benign: malignant = 65:84) enrolled in our study. The median age of them was 44 years (interquartile range, 36–54 years). Among the 149 thyroid nodules, 69 nodules (46%) are located in the left lobe of the thyroid, 78 nodules (52%) are located in the right lobe of the thyroid, and 2 nodules (1%) are located in the isthmus of the thyroid. 25 (17%) of the patients were male and 124 (83%) of them were female. Gender ($p = 0.53$) and location ($p = 1$) were not statistically significant with thyroid cancer while age ($p = 0.02$) was statistically significant with thyroid cancer (**Table1**).

Characteristics of CAD

In the transverse plane of the thyroid, AD ($p = 0.76$) was not statistically significant with thyroid cancer while TD ($p < 0.01$), shape ($p < 0.01$), margin ($p < 0.01$), echogenicity ($p < 0.01$), echogenic foci ($p < 0.01$), composition ($p < 0.01$), TI-RADS classification ($p < 0.01$), and risk ($p < 0.01$) were statistically significant with thyroid cancer (**Table1**).

In the longitudinal plane of the thyroid, AD ($p = 0.85$) was not statistically significant with thyroid cancer while LD ($p = 0.03$), shape ($p < 0.01$), margin ($p < 0.01$), echogenicity ($p < 0.01$), echogenic foci ($p < 0.01$), composition ($p < 0.01$), TI-RADS classification ($p < 0.01$), and risk ($p < 0.01$) were statistically significant with thyroid cancer (**Table1**).

Patients' data distribution between transverse and longitudinal planes is shown in **Figure 3**.

Difference Between Transverse and Longitudinal Planes

PD_{risk}^2 ($p < 0.01$), PD_{shape}^2 ($p < 0.01$), $PD_{echogenic\ foci}^2$($p < 0.01$), and $PD_{composition}^2$($p < 0.01$) were statistically significant with thyroid cancer (**Table2**). PD_{risk} ($p = 0.66$), PD_{shape} ($p = 0.1$), PD_{margin} ($p = 0.05$), PD_{margin}^2 ($p = 0.35$), $PD_{echogenicity}$ ($p = 0.21$), $PD_{echogenicity}^2$ ($p = 0.5$), $PD_{echogenic\ foci}$ ($p = 0.55$), and $PD_{composition}$ ($p = 0.11$) were not statistically significant with thyroid cancer (**Table2**).

Consistency of CAD Features in Transverse and Longitudinal Planes

The ICC of AD, TD, and LD between the ultrasonographer's diagnosis and CAD were 0.97 (0.95–0.98), 0.98 (0.98–0.99), and

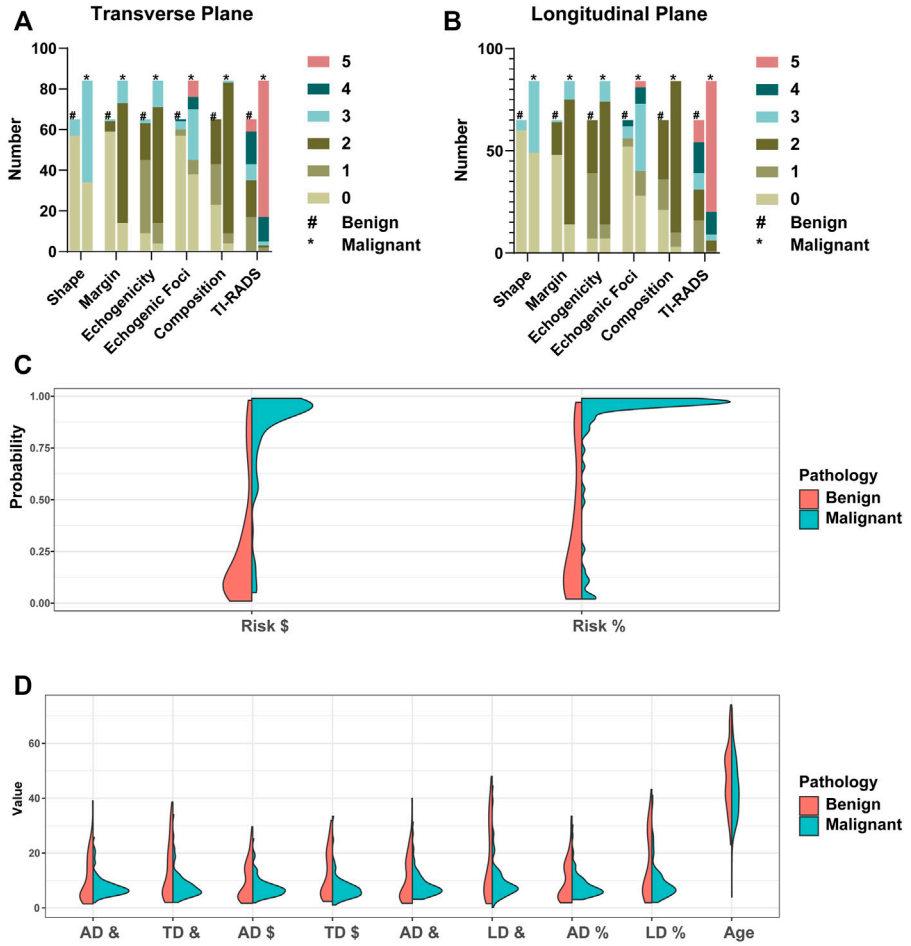


FIGURE 3 | Patients' data distribution between transverse and longitudinal planes. TI-RADS lexicon is shown in **(A,B)**. #Benign group patients' data distribution. *Malignant group patients' data distribution. $Based on the CAD of the transverse plane. %Based on the CAD of the longitudinal plane. &Based on the ultrasonographer's diagnosis. TD: transverse diameter. LD: longitudinal diameter. AD: anteroposterior diameter. The risk score from the CAD system is shown in **(C)**. The TD, LD, AD, and age are shown in **(D)**.

TABLE 2 | Data distribution of PD and PD^2.

Variable	Total (n = 149)	Benign (n = 65)	Malignant (n = 84)	p
PD_{risk}	−0.02 (−0.12, 0.02)	−0.02 (−0.21, 0.04)	−0.02 (−0.06, 0.01)	0.66
PD_{risk}^2	0 (0, 0.04)	0.01 (0, 0.08)	0 (0, 0.01)	<0.01
PD_{shape}	0 (0, 0)	0 (0, 0)	0 (0, 3)	0.1
PD_{shape}^2	0 (0, 9)	0 (0, 0)	0 (0, 9)	<0.01
PD_{margin}	0 (0, 0)	0 (0, 0)	0 (0, 0)	0.05
PD_{margin}^2	0 (0, 1)	0 (0, 0)	0 (0, 1)	0.35
$PD_{echogenicity}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0.21
$PD_{echogenicity}^2$	0 (0, 1)	0 (0, 1)	0 (0, 1)	0.5
$PD_{echogenic\ foci}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0.55
$PD_{echogenic\ foci}^2$	0 (0, 1)	0 (0, 0)	0 (0, 4)	<0.01
$PD_{composition}$	0 (0, 0)	0 (-1, 0)	0 (0, 0)	0.11
$PD_{composition}^2$	0 (0, 0)	0 (0, 1)	0 (0, 0)	<0.01

All the variables are shown by median (first quartile, third quantile), PD: plane difference, and PD^2: plane difference factor.

0.98 (0.97–0.98), respectively, while the ICC of risk between the transverse plane and longitudinal plane of CAD was 0.81 (0.73–0.86), meaning that they have high consistency. The Kappa (mean ± standard error) of TI-RADS classification and shape between the transverse plane and longitudinal plane of CAD are 0.40 ± 0.05 and 0.34 ± 0.08, respectively, meaning that

TABLE 3 | The consistency of CAD features.

	ICC (95% CI)[d]		Kappa[e]
Anteroposterior diameter[a]	0.97 (0.95–0.98)	Shape[c]	0.34 ± 0.08
Transverse diameter[a]	0.98 (0.98–0.99)	Margin[c]	0.47 ± 0.06
Anteroposterior diameter[b]	0.96 (0.94–0.97)	Echogenicity[c]	0.45 ± 0.06
Longitudinal diameter[b]	0.98 (0.97–0.98)	Echogenic foci[c]	0.46 ± 0.06
Risk[c]	0.81 (0.73–0.86)	Composition[c]	0.54 ± 0.06
TI-RADS classification[c]	0.40 ± 0.05		

[a]*The consistency between the ultrasonographer's diagnosis and CAD of the transverse plane.*
[b]*The consistency between the ultrasonographer's diagnosis and CAD of the longitudinal plane.*
[c]*The consistency between the transverse plane and longitudinal plane of CAD.*
[d]*Shown by median (first quartile, third quantile).*
[e]*Shown by mean ± standard error.*
CAD, computer-aided diagnosis; ICC, interclass correlation coefficient; CI: confidence interval.

they have low consistency. The Kappa (mean ± standard error) of margin, echogenicity, echogenic foci, and composition between the transverse plane and longitudinal plane of CAD is 0.47 ± 0.06, 0.45 ± 0.06, 0.46 ± 0.06, and 0.54 ± 0.06, respectively, meaning that they have moderate consistency (**Table 3**).

CAD Features' Diagnosis Efficiency

In two mutually perpendicular planes, **Figure 4A** and **Table 4** demonstrate the ROC for various combinations of TI-RADS classifications and risk. TI-RADS classification and risk had the best diagnostic performance among the 7 features recorded by CAD in mutually perpendicular planes (**Figures 4B–D**).

The AUC (95% CI) of TI-RADS classification in the transverse plane of CAD is better than that of the longitudinal plane [0.90 (0.84–0.95) vs. 0.83 (0.77–0.90), $p = 0.04$]. The AUC (95% CI) of risk in the transverse plane of CAD shows no difference from that in the longitudinal plane statistically [0.90 (0.85–0.95) vs. 0.88 (0.82–0.94), $p = 0.52$].

While combining the CAD features, the diagnosis efficiency will be better. The AUC (95% CI), specificity, sensitivity, and accuracy [TI-RADS classification (transverse plane) + TI-RADS classification (longitudinal plane) + risk (transverse plane) + risk (longitudinal plane)] are 0.93 (0.89–0.97), 86.15%, 90.48%, and 88.59%, respectively. The AUC (95% CI), specificity, sensitivity, and accuracy [TI-RADS classification (transverse plane) + risk (transverse plane)] are 0.91 (0.86–0.96), 86.15%, 89.29%, and 87.92%, respectively. The AUC (95% CI), specificity, sensitivity, and accuracy [TI-RADS classification (longitudinal plane) + risk (longitudinal plane)] are 0.90 (0.85–0.95), 89.23%, 80.95%, and 84.56%, respectively. The AUC (95%CI), specificity, sensitivity, and accuracy [risk (transverse plane) + risk (longitudinal plane)] are 0.92 (0.88–0.97), 84.62%, 89.29%, and 87.25%, respectively. The AUC (95% CI), specificity, sensitivity, and accuracy [TI-RADS classification (transverse plane) + TI-RADS classification (longitudinal plane)] are 0.91 (0.86–0.96), 92.31%, 85.71%, and 88.59%, respectively.

Of the diagnoses made by ultrasonographers of different seniority, the AUC (95% CI), specificity, sensitivity, and accuracy of the 15 years experienced ultrasonographer were 0.90 (0.85–0.95), 81.54%, 89.29%, and 85.91% and that of 10 years experienced ultrasonographer were 0.88 (0.82–0.94), 86.15%, 83.33%, and 84.56% and that of 5 years experienced ultrasonographer were 0.86 (0.80–0.92), 84.62%, 79.76%, and 81.88%, respectively.

DISCUSSION

The incidence of thyroid cancer is increasing year by year. However, the mortality rate of thyroid cancer has not changed (Sosa et al., 2013; Ahn et al., 2016; Leboulleux et al., 2016; Vaccarella et al., 2016). Thus, it is necessary for clinicians to reduce punctures and surgeries for thyroid nodules. ACR TI-RADS, a lexicon for imaging practitioners reporting thyroid nodules, has developed a standardized risk stratification system for thyroid nodules (Tessler et al., 2017; Tessler et al., 2018). Unlike ultrasonographers who are rated according to ACR TI-RADS, S-thyroid is reproducible and objective according to ACR TI-RADS. The objective of our study is to inform clinicians on how to respond when CAD scores different risk scores based on two mutually perpendicular planes so that better clinical protocol decisions can be made that are more beneficial to patients.

We can see that TD, LD, and shape were statistically significant with thyroid cancer while AD was not, and the TD and LD of malignant nodules are smaller than those of benign nodules. Since benign and cystic nodules have softer nodules and less infiltration of surrounding tissue and are therefore more easily compressed than malignant nodules (Yoon et al., 2010), 56%–89% of papillary thyroid cancers showed dense fibrosis (Vickery, 1983; Isarangkul, 1993). This may be the reason why malignant nodules are less likely to be compressed. However, the reasons for the result of our study are more likely due to selection bias. In the case of small nodules, it is likely that only those with suspicious features underwent a biopsy or surgery. Further expansion of the sample is needed to compare the statistical results again.

In our study, we observed that the ACR TI-RADS lexicon diagnosis of CAD based on the transverse plane did differ from that of CAD based on the longitudinal plane, but the difference was not statistically significant with thyroid cancer, while the square of the difference was statistically significant with thyroid. Surprisingly, the correlation of the TI-RADS classification of CAD based on two mutually perpendicular planes was low, but the correlation of risk was high. Therefore, the ultrasonographer or clinician should give priority to the risk score of CAD over the TI-RADS classification of CAD when interpreting CAD reports. Similarly, the correlations of margin, echogenicity, echogenic foci, and composition based on two mutually perpendicular planes of CAD were moderate. This may be the reason for the low correlation of the TI-RADS classification based on two mutually perpendicular planes in our study. We know that the images of thyroid nodules are not consistent in both planes, so ultrasonographers base their grading of thyroid nodules on the combined scan. This is different from the TI-RADS classification of CAD, which is why the previous results occur. However, this phenomenon has not

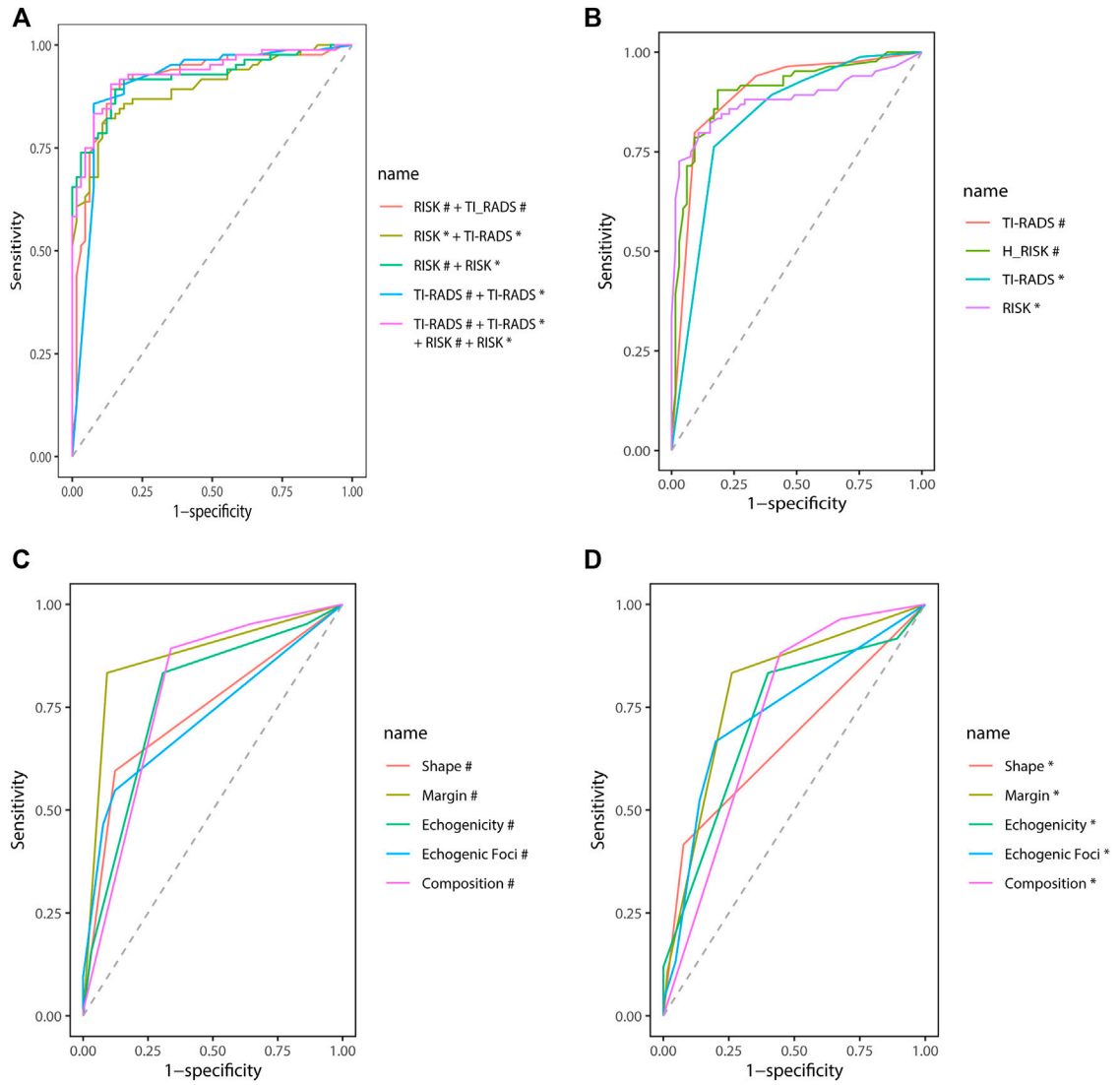


FIGURE 4 | Receiver operating characteristic (ROC). #Based on the CAD of the transverse plane. *Based on the CAD of the longitudinal plane.

TABLE 4 | ROC of different combination of CAD features' diagnosis efficiency.

	AUC (95%CI)	Threshold	Specificity	Sensitivity	Accuracy
TI-RADS[c] + TI-RADS[d] + risk[c] + risk[d]	0.93 (0.89–0.97)	0.58	86.15	90.48	88.59
Risk[c] + TI_RADS[c]	0.91 (0.86–0.96)	0.65	86.15	89.29	87.92
Risk[d] + TI-RADS[d]	0.90 (0.85–0.95)	0.77	89.23	80.95	84.56
Risk[c] + risk[d]	0.92 (0.88–0.97)	0.57	84.62	89.29	87.25
TI-RADS[c] + TI-RADS[d]	0.91 (0.86–0.96)	0.66	92.31	85.71	88.59
TI-RADS[c]	0.90 (0.84–0.95)	4.5	90.77	79.76	84.56
Risk[c]	0.90 (0.85–0.95)	0.5	81.54	90.48	86.58
TI-RADS[d]	0.83 (0.77–0.90)	4.5	83.08	76.19	79.19
Risk[d]	0.88 (0.82–0.94)	0.94	96.92	72.62	83.22

[a]*Based on the CAD of the transverse plane.*
[b]*Based on the CAD of the longitudinal plane.*
[c]*Non-normally distributed numerical variables are shown by median (first quartile, third quantile).*
[d]*Disordered classification variables are shown by percentage.*

been analyzed before, so what should an ultrasonographer or clinician do when interpreting CAD results that are inconsistent based on two mutually perpendicular planes?

Therefore, this study also investigated the diagnostic efficacy of TI-RADS classification and risk in CAD transverse and longitudinal planes. It was found that the diagnosis of thyroid cancer in the CAD transverse plane was superior to the CAD longitudinal plane when using the TI-RADS classification, but there was no difference in the diagnosis between the two planes when using risk. What is more, the combination of both planes can improve the diagnosis of thyroid cancer. Therefore, the ultrasonographer or clinician should not interpret the CAD results based on one plane alone but should combine the results of both the transverse and longitudinal planes.

Wei Q et al. found that S-detect can improve diagnostic performance for less experienced radiologists, and the sensitivity and specificity of S-detect are 91.3% and 65.2%, respectively (Wei et al., 2020). Kim HL et al. evaluated the diagnostic performance of S-Detect 1 and S-Detect 2 for detecting thyroid cancers and found that the sensitivity and specificity of S-Detect 1 are 80.2% and 82.6% and that of S-Detect 2 are 81.4% and 68.2%, respectively (Kim et al., 2019). Xia S et al. found that CAD presents a higher sensitivity but lower specificity in identifying malignant thyroid nodules compared to experienced radiologists, and the sensitivity and specificity of S-detect are 90.5% and 41.2%, respectively (Xia et al., 2019). Barczyński M et al. found that the CAD system has similar sensitivity to classify thyroid lesions as a surgeon with expert US skills (Barczyński et al., 2020). In conclusion, all the aforementioned studies demonstrated the high sensitivity and low specificity of S-detect in the diagnosis of malignant thyroid nodules. However, in our study, S-thyroid showed high sensitivity and specificity for identifying thyroid cancer with a combination of transverse and longitudinal planes, 90.48% and 86.15%, respectively. Interestingly, S-thyroid results in either transverse or longitudinal planes alone have low sensitivity and high specificity in the identification of thyroid cancer. The reason for this result may be due to the different algorithms of S-thyroid and S-detect, or it may be due to the fact that the study sample of this study is different from the study sample of the previous study, and further comparison of the two CAD software with the same patient sample is needed.

There are several limitations to our study. First, we selected patients who had undergone surgery, so the ratio of benign to malignant thyroid nodules was not correct, and there were more malignant nodules than benign ones, which may affect the diagnostic performance of the CAD system. Second, non-mass lesions were not included in the study population because the CAD analysis was limited to non-mass lesions. Last but not the least, the number of patient cases in this study was too small, and further sample size studies are needed.

CONCLUSION

In our study, we explored the diagnostic capability of S-thyroid, using CAD software for thyroid nodules based on two mutually perpendicular planes and found that the best diagnostic capability was achieved with a combination of CAD transverse and longitudinal planes.

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.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Shenzhen People's 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.

FUNDING

This project was supported by the Commission of Science and Technology of Shenzhen (JCYJ20190806151807192). Thanks to our D. team of the Ultrasound Department of Shenzhen People's Hospital.

REFERENCES

Ahn, H. S., Kim, H. J., Kim, K. H., Lee, Y. S., Han, S. J., Kim, Y., et al. (2016). Thyroid Cancer Screening in South Korea Increases Detection of Papillary Cancers With No Impact on Other Subtypes or Thyroid Cancer Mortality. *Thyroid* 26, 1535–1540. doi:10.1089/thy.2016.0075

Barczyński, M., Stopa-Barczyńska, M., Wojtczak, B., Czarniecka, A., and Konturek, A. (2020). Clinical Validation of S-DetectTM Mode in Semi-Automated Ultrasound Classification of Thyroid Lesions in Surgical Office. *Gland Surg.* 9, S77–s85. doi:10.21037/gs.2019.12.23

Choi, Y. J., Baek, J. H., Park, H. S., Shim, W. H., Kim, T. Y., Shong, Y. K., et al. (2017). A Computer-Aided Diagnosis System Using Artificial Intelligence for the Diagnosis and Characterization of Thyroid Nodules on Ultrasound: Initial Clinical Assessment. *Thyroid* 27, 546–552. doi:10.1089/thy.2016.0372

Filetti, S., Durante, C., Hartl, D., Leboulleux, S., Locati, L. D., Newbold, K., et al. (2019). Thyroid Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* 30, 1856–1883. doi:10.1093/annonc/mdz400

Grani, G., and Fumarola, A. (2014). Thyroglobulin in Lymph Node fine-needle Aspiration Washout: a Systematic Review and Meta-Analysis of Diagnostic Accuracy. *J. Clin. Endocrinol. Metab.* 99, 1970–1982. doi:10.1210/jc.2014-1098

Grani, G., Sponziello, M., Pecce, V., Ramundo, V., and Durante, C. (2020). Contemporary Thyroid Nodule Evaluation and Management. *J. Clin. Endocrinol. Metab.* 105, 2869–2883. doi:10.1210/clinem/dgaa322

Hoang, J. K., Middleton, W. D., Farjat, A. E., Teefey, S. A., Abinanti, N., Boschini, F. J., et al. (2018). Interobserver Variability of Sonographic Features Used in the American College of Radiology Thyroid Imaging Reporting and Data System. *Am. J. Roentgenology* 211, 162–167. doi:10.2214/ajr.17.19192

Isarangkul, W. (1993). Dense Fibrosis. Another Diagnostic Criterion for Papillary Thyroid Carcinoma. *Arch. Pathol. Lab. Med.* 117, 645–646.

Kim, H. L., Ha, E. J., and Han, M. (2019). Real-World Performance of Computer-Aided Diagnosis System for Thyroid Nodules Using Ultrasonography. *Ultrasound Med. Biol.* 45, 2672–2678. doi:10.1016/j.ultrasmedbio.2019.05.032

Leboulleux, S., Tuttle, R. M., Pacini, F., and Schlumberger, M. (2016). Papillary Thyroid Microcarcinoma: Time to Shift from Surgery to Active Surveillance? *Lancet Diabetes Endocrinol.* 4, 933–942. doi:10.1016/s2213-8587(16)30180-2

Lee, Y. J., Choi, S. Y., Kim, K. S., and Yang, P. S. (2016). Variability in Observer Performance between Faculty Members and Residents Using Breast Imaging Reporting and Data System (BI-RADS)-Ultrasound, Fifth Edition (2013). *Iran J. Radiol.* 13, e28281. doi:10.5812/iranjradiol.28281

Nambron, R., Rosenthal, R., and Bahl, D. (2020). Diagnosis and Evaluation of Thyroid Nodules-The Clinician's Perspective. *Radiologic Clin. North America* 58, 1009–1018. doi:10.1016/j.rcl.2020.07.007

Persichetti, A., Di Stasio, E., Coccaro, C., Graziano, F., Bianchini, A., Di Donna, V., et al. (2020). Inter- and Intraobserver Agreement in the Assessment of Thyroid Nodule Ultrasound Features and Classification Systems: A Blinded Multicenter Study. *Thyroid* 30, 237–242. doi:10.1089/thy.2019.0360

Shen, J., Zhang, C. J. P., Jiang, B., Chen, J., Song, J., Liu, Z., et al. (2019). Artificial Intelligence versus Clinicians in Disease Diagnosis: Systematic Review. *JMIR Med. Inform.* 7, e10010. doi:10.2196/10010

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics, 2021. *CA A. Cancer J. Clin.* 71, 7–33. doi:10.3322/caac.21654

Singh, S., Maxwell, J., Baker, J. A., Nicholas, J. L., and Lo, J. Y. (2011). Computer-aided Classification of Breast Masses: Performance and Interobserver Variability of Expert Radiologists versus Residents. *Radiology* 258, 73–80. doi:10.1148/radiol.10081308

Sosa, J. A., Hanna, J. W., Robinson, K. A., and Lanman, R. B. (2013). Increases in Thyroid Nodule fine-needle Aspirations, Operations, and Diagnoses of Thyroid Cancer in the United States. *Surgery* 154, 1420–1427. doi:10.1016/j.surg.2013.07.006

Tessler, F. N., Middleton, W. D., Grant, E. G., Hoang, J. K., Berland, L. L., Teefey, S. A., et al. (2017). ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J. Am. Coll. Radiol.* 14, 587–595. doi:10.1016/j.jacr.2017.01.046

Tessler, F. N., Middleton, W. D., and Grant, E. G. (2018). Thyroid Imaging Reporting and Data System (TI-RADS): A User's Guide. *Radiology* 287, 29–36. doi:10.1148/radiol.2017171240

Vaccarella, S., Franceschi, S., Bray, F., Wild, C. P., Plummer, M., and Dal Maso, L. (2016). Worldwide Thyroid-Cancer Epidemic? the Increasing Impact of Overdiagnosis. *N. Engl. J. Med.* 375, 614–617. doi:10.1056/nejmp1604412

Vickery, A. L., Jr. (1983). Thyroid Papillary Carcinoma. *Am. J. Surg. Pathol.* 7, 797–807. doi:10.1097/00000478-198307080-00009

Wei, Q., Zeng, S.-E., Wang, L.-P., Yan, Y.-J., Wang, T., Xu, J.-W., et al. (2020). The Value of S-Detect in Improving the Diagnostic Performance of Radiologists for the Differential Diagnosis of Thyroid Nodules. *Med. Ultrason.* 22, 415–423. doi:10.11152/mu-2501

Xia, S., Yao, J., Zhou, W., Dong, Y., Xu, S., Zhou, J., et al. (2019). A Computer-Aided Diagnosing System in the Evaluation of Thyroid Nodules-Experience in a Specialized Thyroid center. *World J. Surg. Onc* 17, 210. doi:10.1186/s12957-019-1752-z

Yoon, S. J., Yoon, D. Y., Chang, S. K., Seo, Y. L., Yun, E. J., Choi, C. S., et al. (2010). "Taller-Than-Wide Sign" of Thyroid Malignancy: Comparison between Ultrasound and CT. *Am. J. Roentgenology* 194, W420–W424. doi:10.2214/ajr.09.3376

Zhang, D., Jiang, F., Yin, R., Wu, G. G., Wei, Q., Cui, X. W., et al. (2021). A Review of the Role of the S-Detect Computer-Aided Diagnostic Ultrasound System in the Evaluation of Benign and Malignant Breast and Thyroid Masses. *Med. Sci. Monit.* 27, e931957. doi:10.12659/msm.931957

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, Chen, Wu, Tian, Ye, Xu, Luo and Dong. 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 Physiology

ORIGINAL RESEARCH
published: 24 May 2022
doi: 10.3389/fphys.2022.870657

Construction of an HLA Classifier for Early Diagnosis, Prognosis, and Recognition of Immunosuppression in Sepsis by Multiple Transcriptome Datasets

Zhen Chen[1†], Rui Chen[2†], Yangpeng Ou[3†], Jianhai Lu[1†], Qianhua Jiang[1†], Genglong Liu[4], Liping Wang[5], Yayun Liu[6], Zhujiang Zhou[1], Ben Yang[7] and Liuer Zuo[1*†]*

[1]Department of Intensive Care Unit, Shunde Hospital, Southern Medical University (The First People's Hospital of Shunde), Foshan, China, [2]Department of Medical Intensive Care Unit, General Hospital of Southern Theater Command, Guangzhou, China, [3]Department of Oncology, Huizhou Third People's Hospital, Guangzhou Medical University, Huizhou, China, [4]Department of Pathology, The Third Affiliated Hospital of Guangdong Medical University (Longjiang Hospital of Shunde District), Foshan, China, [5]Department of Otorhinolaryngology Head and Neck Surgery, The First Affiliated Hospital of Hainan Medical University, Haikou, China, [6]Department of Endocrinology, GuiYang Huaxi District People's Hospital, Guiyang, China, [7]Department of Burn Surgery, Huizhou Municipal Central Hospital, Huizhou, China

Background: Sepsis is a clinical syndrome, due to a dysregulated inflammatory response to infection. Accumulating evidence shows that human leukocyte antigen (HLA) genes play a key role in the immune responses to sepsis. Nevertheless, the effects of HLA genes in sepsis have still not been comprehensively understood.

Methods: A systematical search was performed in the Gene Expression Omnibus (GEO) and ArrayExpress databases from inception to 10 September 2021. Random forest (RF) and modified Lasso penalized regression were conducted to identify hub genes in multi-transcriptome data, thus we constructed a prediction model, namely the HLA classifier. ArrayExpress databases, as external validation, were utilized to evaluate its diagnostic, prognostic, and predictive performance. Immune cell infiltration score was calculated via CIBERSORTx tools and single-sample gene set enrichment analysis (ssGSEA). Gene set variation analysis (GSVA) and ssGSEA were conducted to determine the pathways that are significantly enriched in different subgroups. Next, we systematically correlated the HLA classifier with immunological characteristics from multiple perspectives, such as immune-related cell infiltration, pivotal molecular pathways, and cytokine expression. Finally, quantitative real-time polymerase chain reaction (qRT-PCR) was conducted to validate the expression level of HLA genes in clinical samples.

OPEN ACCESS

Edited by:
Wei Wang,
Huzhou Maternity and Child Health Care Hospital, China

Reviewed by:
Kanhaiya Singh,
Indiana University and Purdue University Indianapolis, United States
Oscar Medina-Contreras,
Federico Gómez Children's Hospital, Mexico

***Correspondence:**
Zhen Chen
jeanyz@foxmail.com
Liuer Zuo
13500276597@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

Received: *22 February 2022*
Accepted: *18 April 2022*
Published: *24 May 2022*

Citation:
Chen Z, Chen R, Ou Y, Lu J, Jiang Q, Liu G, Wang L, Liu Y, Zhou Z, Yang B and Zuo L (2022) Construction of an HLA Classifier for Early Diagnosis, Prognosis, and Recognition of Immunosuppression in Sepsis by Multiple Transcriptome Datasets. Front. Physiol. 13:870657. doi: 10.3389/fphys.2022.870657

Abbreviations: APACHE II, acute physiology and chronic health evaluation; APC, antigen-presenting cell; CCR, cytokine-cytokine receptor interaction; CRP, C-reactive protein; DCA, decision curve analysis; DM, diabetes mellitus; GEO, gene expression omnibus; GSVA, gene set variation analysis; HLA, human leukocyte antigen; IL, infiltrating lymphocyte; MARS, molecular diagnosis and risk stratification of sepsis; MDSCs, myeloid-derived suppressor cells; mHLA-DR, monocyte HLA-DR; MSigDB, molecular signatures Database; PCT, procalcitonin; qSOFA, quick sequential organ failure; RF, random forest; ROC, receiver operating characteristic; SSC, surviving sepsis campaign; ssGSEA, single-sample gene set enrichment analysis; SRS, sepsis response signature.

Results: A total of nine datasets comprising 1,251 patients were included. Based on RF and modified Lasso penalized regression in multi-transcriptome datasets, five HLA genes (B2M, HLA-DQA1, HLA-DPA1, TAP1, and TAP2) were identified as hub genes, which were used to construct an HLA classifier. In the discovery cohort, the HLA classifier exhibited superior diagnostic value (AUC = 0.997) and performed better in predicting mortality (AUC = 0.716) than clinical characteristics or endotypes. Encouragingly, similar results were observed in the ArrayExpress databases. In the E-MTAB-7581 dataset, the use of hydrocortisone in the HLA high-risk subgroup (OR: 2.84, 95% CI 1.07–7.57, p = 0.037) was associated with increased risk of mortality, but not in the HLA low-risk subgroup. Additionally, immune infiltration analysis by CIBERSORTx and ssGSEA revealed that B cells, activated dendritic cells, NK cells, T helper cells, and infiltrating lymphocytes (ILs) were significantly richer in HLA low-risk phenotypes, while Tregs and myeloid-derived suppressor cells (MDSCs) were more abundant in HLA high-risk phenotypes. The HLA classifier was significantly negatively correlated with B cells, activated dendritic cells, NK cells, T helper cells, and ILs, yet was significantly positively correlated with Tregs and MDSCs. Subsequently, molecular pathways analysis uncovered that cytokine-cytokine receptor (CCR) interaction, human leukocyte antigen (HLA), and antigen-presenting cell (APC) co-stimulation were significantly enriched in HLA low-risk endotypes, which was significantly negatively correlated with the HLA classifier in multi-transcriptome data. Finally, the expression levels of several cytokines (IL-10, IFNG, TNF) were significantly different between the HLA subgroups, and the ratio of IL-10/TNF was significantly positively correlated with HLA score in multi-transcriptome data. Results of qRT-PCR validated the higher expression level of B2M as well as lower expression level of HLA-DQA1, HLA-DPA1, TAP1, and TAP2 in sepsis samples compared to control sample.

Conclusion: Based on five HLA genes, a diagnostic and prognostic model, namely the HLA classifier, was established, which is closely correlated with responses to hydrocortisone and immunosuppression status and might facilitate personalized counseling for specific therapy.

Keywords: sepsis, HLA genes, immune infiltration, immunosuppression, model

KEY MESSAGES

1) To the best of our knowledge, this is the first comprehensive study to explore the HLA family based on multiple transcriptome expression profiles in all-cause sepsis, leading to the discovery of novel biomarkers to develop a diagnostic and prognostic model, thus elucidating the model and immune system (immune cells infiltration, immune-related pathways, and cytokines) to find its additional clinical implications.
2) Based on random forest and modified Lasso penalized regression, a diagnostic and prognostic model (HLA classifier) was constructed, which could be a robust tool to diagnose sepsis earlier and to identify patients at risk of a poor or even fatal outcome. Additionally, the HLA classifier is closely correlated with responses to hydrocortisone and may be useful for clinicians to tailor treatment decisions. According to immune cell infiltration, immune-related pathways, and cytokines level, the HLA classifier could efficiently reflect immunological status, which may help guide immune-modulating agents to achieve immune homeostasis.
3) Results of qRT-PCR validated a higher expression level of B2M as well as a lower expression level of HLA-DQA1, HLA-DPA1, TAP1, and TAP2 in sepsis samples compared to control samples, which were in accordance with the results of bioinformatics analyses derived from the GEO datasets.

INTRODUCTION

Sepsis, a life-threatening syndrome characterized by organ failure after infection, is caused by a dysregulated host response to infection (Singer et al., 2016). Clinical epidemiological analyses show that the estimated national cases and in-hospital mortality cases of sepsis were approximately 48.9 million and 11.0 million,

respectively, representing one-fifth of all causes of death, and making it one of the major socioeconomic burdens all over the world (Rudd et al., 2020). In the past decade, according to the recommendations of the Surviving Sepsis Campaign (SSC) for the management of sepsis patients, the mortality rate decreased from about 37%–25%, whereas this figure is still too high to be acceptable (Fleischmann et al., 2016). To fight the global burden of sepsis, given the lack of obvious and nonspecific clinical signs in early-stage disease, early diagnosis and appropriate treatment is critical to improve patients' outcomes, on account of the fact that each hour of delay in initiating treatment is associated with increased mortality rates (Ferrer et al., 2014). Importantly, classification and identification of a patient at high risk may aid clinicians to screen and identify individuals who are most likely to benefit from additional monitoring and treatment, or to detect an immunosuppressed state which could benefit from targeted immunostimulating therapies, and eventually improve patient prognosis.

As sepsis is a highly intricate condition and its clinical evaluation is often challenging, the additional usage of biomarkers for rapid diagnosis that help pinpoint high-risk patients is an attractive solution. Currently, several biomarkers, such as C-reactive protein (CRP), which is characterized as a inflammatory marker, and procalcitonin (PCT), which serves as a marker of bacteremia, have been widely utilized as acute phase reactants in critically ill patients, yet their diagnostic and prognostic performance for sepsis are suboptimal (Henriquez-Camacho and Losa, 2014). Recently, the quick sequential organ failure (qSOFA) score was introduced as a bedside standard based on three clinical elements, which was generated through a data-driven approach. However, it has been controversial since it was proposed, in terms of the diagnosis, it has high sensitivity and low specificity, yet with regard to the prognosis, it has low sensitivity and high specificity, which makes its implementation problematic (Simpson, 2017). To date, none of the signatures of the immune response or circulating blood biomarkers that have been investigated detect sepsis quickly enough or recognize high-risk patients with an acceptable certainty, which was ascribed to the heterogeneity and complex pathophysiology of sepsis. To a certain extent, the heterogeneity can be related to the differential expression of thousands of genes in response to infectious stimuli (Zhai et al., 2020). Hence, transcriptomics, as promising new biomarkers, can provide important predictive and prognostic information.

Pathophysiologically, human leukocyte antigen (HLA), cross-link innate and adaptive, plays an important role in recognition, processing, and presentation of protein antigens (such as organisms) to cognate T cells, NK cells, etc., therefore, initiating an immune response, which is involved in the pathogenesis of sepsis (Koşaloğlu-Yalçın et al., 2018). HLA-DR, which belongs to HLA class II and is located on chromosome 6, plays a crucial role in modulating immune responses. Diminished monocyte HLA-DR (mHLA-DR) expression on the cell surface was associated with increased risk of death of septic patients or adverse outcomes and increased susceptibility to nosocomial/secondary infections (Hotchkiss et al., 2016). It is noteworthy that mHLA-DR is considered an effective indicator of the general immunoparalysis state of patients. However, on account of inferior specificity and sensitivity, the capacity of HLA-DR expression monitoring to predict mortality is not completely recognized (Monneret et al., 2006). Currently, no single biomarker can be useful for diagnosing sepsis, prognosis, and disease monitoring due to patient heterogeneity. A panel of markers ("sepsis signature") seems to be able to offer better predictive and prognostic purposes. The availability of genome-sequencing data opened up the possibility of comprehensively investigating all HLA gene alterations of sepsis, resulting in the discovery of new biomarkers for early identification, disease surveillance, and to guide specific adjuvant therapy.

A comprehensive characterization of the HLA family in the adult host response to all-cause sepsis has not previously been done. In our current study, based on random forest (RF) and modified Lasso penalized regression, we identified hub HLA genes, and constructed a prediction model, namely the HLA classifier. Subsequently, the predictive and prognostic values of the model were tested in independent validation cohorts from ArrayExpress databases. Finally, we systematically correlated the HLA classifier with immunological characteristics from multiple perspectives, such as immune-related cell infiltration, pivotal molecular pathways, and cytokine expression.

MATERIALS AND METHODS

Sample Selection, Data Acquisition, and Processing

A comprehensive search was performed in Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) and ArrayExpress (http://www.ncbi.nlm.nih.gov/geo) databases from 9 July 2005 to 10 September 2021 to identify relevant transcriptomic profiling datasets. The inclusion criteria were the following: expression profiling by array or high-throughput sequencing: sepsis; organism: homo sapiens; samples size more than 50; adult patients (more than 18 years old). Ultimately, six GEO datasets, as discovery cohorts, and three ArrayExpress databases, as external validation cohorts, fulfilled our eligibility criteria and were included for both qualitative and quantitative analysis. The basic information of these microarray datasets is listed in **Table 1**. Additionally, a panel of 28 HLA genes was collected from published studies (Schaafsma et al., 2021) (**Supplementary Table S1**). All data were normalized with the edgeR package or Limma package in the R computing environment.

Clinical Specimens

Peripheral blood mononuclear cells (PBMCs) were collected from 50 clinical samples, including 25 sepsis samples and 25 healthy controls. The study was reviewed and approved by the institutional review board (Ethics Committee) of the Shunde

TABLE 1 | Dataset included in the study.

Accession	Cohort description	Timing of gene expression profiling	Country	Normal sample	Mortality/ sepsis sample	Sample type
GSE65682	Sepsis due to CAP and HAP + AS	On ICU admission	Netherlands and United Kingdom	42	48/231	Whole blood
GSE54514	Sepsis	Within 24 h of ICU admission	Australia	18	9/35	Whole blood
GSE57065	Septic shock	On ICU admission	France	25	-/28	Whole blood
GSE63042	Sepsis	Day of enrollment upon presentation to the ED	United States	—	28/106	Whole blood
GSE69528	Sepsis due to CAP	On ICU admission	United States	55	-/83	Whole blood
GSE95233	Septic shock	Day 1 of ICU admission	France	22	34/51	Whole blood
E-MTAB-4421	Septic shock	On ICU admission	United Kingdom	—	56/265	Whole blood
E-MTAB-4451	Sepsis due to CAP	On ICU admission	United Kingdom	—	57/114	Whole blood
E-MTAB-7581	Septic shock	At enrollment	United Kingdom	—	48/176	Whole blood

Abbreviations: CAP, community acquired pneumonia; HAP, hospital acquired pneumonia; AS, abdominal sepsis; ICU, intensive care medicine; ED, emergency room.

Hospital, Southern Medical University (the First People's Hospital of Shunde).

Identification of Hub HLA Genes and Construction of an HLA Classifier

To select out convincing hub HLA genes, machine learning approaches, including modified Lasso penalized regression and RF (random forest), were adopted. A Lasso regression was performed with 10-fold cross-validation to identify candidate HLA genes and was run for 1,000 cycles to select feature variables based on minimum criteria or 1—s.e. criteria. RF (random forest), a tree-based ensemble comprised of tree-structured classifiers, was established to select feature variables via the package "randomForest" with minimum error regression trees. The importance of variables was ranked using IncNodePurity. The real hub HLA genes were obtained from the intersection of the result of Lasso and RF (GSE65682, GSE63042, and GSE95233 datasets), which was used to develop a prediction model, namely the HLA classifier. The HLA score was generated through a linear combination of coefficients from logistic regression and the relative expression of each HLA. According to this formula, each patient's HLA score was calculated, and patients were classified into low-risk or high-risk groups on the basis of the optimal cut-off value with the maximal sensitivity and specificity in a receiver operating characteristic (ROC) curve.

Diagnostic and Prognostic Value of the HLA Classifier

ROC analysis using the pROC package was carried out to evaluate diagnostic performance with sepsis as the endpoint. Regarding a prognostic aspect, first univariate and multivariate logistic regression analyses were utilized to adjudicate whether the predictive ability of the HLA classifier remained independent of other clinical features [including age, sex, diabetes mellitus (DM), sepsis response signature (SRS), the Molecular Diagnosis and Risk Stratification of Sepsis (MARS), and Acute Physiology and Chronic Health Evaluation (APACHE II)] in multiple datasets. Then the prognostic value of the HLA classifier was compared against age, SRS, MARS, and APACHE II in the discovery and external validation cohorts.

Clinical Usefulness of the HLA Classifier

We explored whether there was an interaction between the HLA classifier and the treatment (vasopressin versus norepinephrine; hydrocortisone versus placebo) in logistic regression models, by using the binary mortality outcome as the response variable in the E-MTAB-7581 dataset. Additionally, to evaluate the clinical value of the HLA classifier, decision curve analysis (DCA), calculating the net benefit for a range of threshold probabilities which places benefits and harms on the same scale (Vickers and Elkin, 2006), was utilized to compare age, SRS, MARS, and APACHE II in the discovery and external validation cohorts.

Evaluation of Immune Cell Infiltration by CIBERSORTx and ssGSEA

To evaluate relative abundance of immune infiltrates, CIBERSORTx (https://cibersort.stanford.edu/) (Newman et al., 2015), which transforms the normalized gene expression matrix into the composition of infiltrating immune cells, and is a kind of deconvolution algorithm that utilizes 1,000 iterations, was used. We filtered out samples with a CIBERSORTx output of a p-value more than 0.05 for accurate forecast of immune cell composition. The ggplot2 package was used to generate bar graphs visualizing the content of 22 types of infiltrating immune cells in each sample and violin plots were used to display variance analysis of immune cells between HLA subgroups.

The GSVA package in R was used to conduct ssGSEA on a metagene set of 26 immune cell subtypes (**Supplementary Table S2**) that are representative of specific immune cells (Bindea et al., 2013). To determine differential immune cell subtypes between the two subgroups (p-value < 0.05), the two-tailed Wilcoxon test

was utilized to analyze the immunoscores, and violin plots were used to visualize the results. Additionally, we explored the correlation between the HLA classifier and immune cells by Spearman correlation analyses in multiple transcriptome datasets. A $p < 0.05$ would be considered statistically significant.

Immune and Molecular Function Between the HLA Subgroups by GSVA and ssGSEA

GSVA, which converts genes from a sample matrix into predefined gene sets without a priori knowledge of experiment design, is a non-parametric unsupervised approach. The KEGG gene sets (c2.cp.kegg.v7.4.symbols.gmt), which were downloaded from the Molecular Signatures Database (MSigDB) (http://software.broadinstitute.org/gsea/index.jsp) (Hänzelmann et al., 2013), were used to estimate variation of pathway activity in each sample. The significantly enriched pathways in KEGG gene sets were set at p-value < 0.05 and enrichment score change >1.0. Additionally, ssGSEA, which generates an enrichment score to signify the levels of absolute enrichment of a metagene set within certain gene signatures in each sample, was applied to evaluate the enrichment degree of immune-related pathways (Bindea et al., 2013) in current immunology research. **Supplementary Table S3** lists the metagene set. Additionally, we explored the relationship between the HLA classifier and pivotal molecular pathways by Spearman correlation analyses in multiple transcriptome datasets. A $p<0.05$ would be considered statistically significant.

Analyses of the Cytokines

A panel of 27 clinically detectable inflammatory cytokines was collected from published studies (Dong et al., 2019). To further define cytokine expression between the HLA subgroups, two-tailed variance analysis was conducted. Additionally, we explored the relationship between the HLA classifier and cytokine expression level by Spearman correlation analyses in multiple transcriptome datasets. A $p<0.05$ would be considered statistically significant.

Quantitative Real-Time PCR (qRT-PCR)

Following the manufacturer's protocol, Trizol (Invitrogen) was used to extract total RNA from peripheral blood mononuclear cells (PBMCs). Reverse transcription of RNA was completed using a RevertAid RT Reverse Transcription Kit (Thermo Scientific). Quantitative PCR was performed using a PowerUp™ SYBR™ Green Master Mix (Thermo Scientific). The results were standardized with GAPDH. Quantitative reverse transcription PCR was conducted using the ABI 7500 real-time PCR system (Applied Biosystems, Foster City, CA, United States). Fold change was determined as $2^{-\triangle\triangle Ct}$ in gene expression. Gene-specific PCR primers are listed in **Supplementary Table S4**.

Statistical Analysis

R software (R version 3.6.1) was utilized to conduct the statistical analysis. Statistical significance was set at a two-sided $p<0.05$ except for where a certain p-value has been given.

RESULTS

Identification of Hub HLA Genes and Construction of an HLA Classifier

The flow chart of the dataset selection procedure is shown in **Supplementary Figure S1**.

Modified Lasso penalized regression was established to shrink and select out hub HLA genes in the discovery cohort, as shown in **Figures 1A,B** (GSE65682 set), in **Supplementary Figures S2A,B** (GSE69528 set), and in **Supplementary Figures S3A,B** (GSE95233 set). Likewise, an RF was also built with minimum error regression trees for hub HLA genes in the discovery cohort, as displayed in **Figures 1C,D** (GSE65682 set), in **Supplementary Figures S2C,D** (GSE69528 set), and in **Supplementary Figures S3C,D** (GSE95233 set). According to the result of Lasso regression and RF in the discovery cohort, we took the intersection of 6 results to acquire 5 hub genes (B2M, HLA-DQA1, HLA-DPA1, TAP1, and TAP2) shared by ≥ 4 results (**Figure 1E**). Additionally, **Supplementary Figure S4** displays the five hub HLA genes on the location of chromosomes. Finally, the five hub HLA genes were used to develop a prediction model, namely the HLA classifier, and the HLA score was computed.

Diagnostic and Prognostic Value of the HLA Classifier

As displayed in **Figure 2**, the diagnostic ability of the HLA classifier to distinguish sepsis from the control samples showed superior diagnostic efficiency, with an AUC of 0.997 in the GSE65682 datasets, an AUC of 0.966 in the GSE57065 datasets, an AUC of 0.956 in the GSE69528 datasets, and an AUC of 1 in the GSE95233 datasets. However, the diagnostic ability of the housekeeping gene panel displayed inferior diagnostic efficiency for sepsis, with an AUC less than 0.6. As for prognostic value, univariate and multivariate Cox regression analysis confirmed that the HLA score was an independent predictor of unfavorable survival outcome, regardless of other clinical characteristics, in multiple transcriptome datasets (**Table 2**). In addition, as shown in **Figure 3**, ROC analysis was performed to investigate the prognostic value of the HLA classifier in the discovery cohorts, with an AUC of 0.716 in the GSE65682 datasets, an AUC of 0.807 in the GSE63042 datasets, an AUC of 0.813 in the GSE95233 datasets, and an AUC of 1 in the GSE54514 datasets. Similarly, the HLA classifier showed a favorable prognostic ability in the external validation cohort, with an AUC of 0.752 in the E-MTAB-4421 datasets, an AUC of 0.691 in the E-MTAB-4451 datasets, and an AUC of 0.737 in the E-MTAB-7851 datasets (**Supplementary Figure S5**). Based on the optimal cut-off value from the ROC curve, patients were categorized into the low-risk group (n = 133) or high-risk group (n = 98) in the GSE65682 sets, low-risk group (n = 21) or high-risk group (n = 14) in the GSE54514 sets, low-risk group (n = 58) or high-risk group (n = 48) in the GSE63042 sets, low-risk group (n = 25) or high-risk group (n = 26) in the GSE95233 sets, low-risk group (n = 140) or high-risk group (n = 125) in the E-MTAB-4421 sets, low-risk group (n = 29) or high-risk group (n = 77) in the E-MTAB-4451 sets, and low-risk group (n = 69) or high-risk

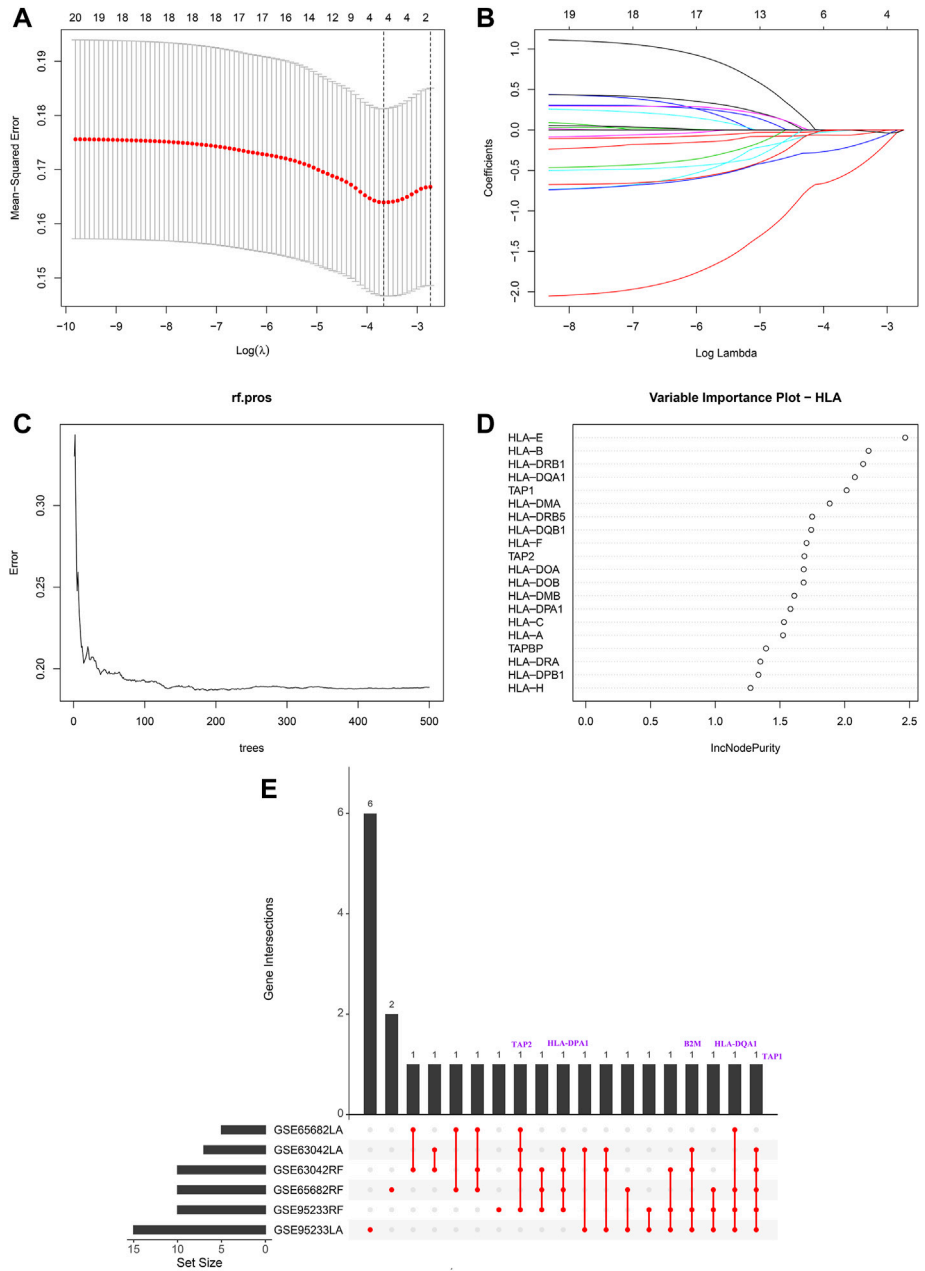


FIGURE 1 | Hub HLA genes selected by Lasso regression analysis and random forest (RF) in GSE65682 datasets. **(A)** The two dotted vertical lines are drawn at the optimal values by minimum criteria (left) and 1—s.e. criteria (right). **(B)** Lasso coefficient profiles of the four hub HLA genes. A vertical line is drawn at the optimal value by minimum criteria and results in four non-zero coefficients. **(C)** Distribution diagram of regression tree and error. **(D)** Important variables ranked by IncNodePurity. **(E)** UpSet plot presents the intersection of six results to identify hub HLA genes.

group (n = 107) in the E-MTAB-4421 sets. Patients in the high-risk group showed a significantly higher mortality rate than in the low-risk group (p <0.001 for Chi-square test) in multiple transcriptome datasets (**Figure 3** and **Supplementary Figure S5**). Importantly, the HLA classifier (AUC: 0.716) performed better in predicting mortality than age (AUC: 0.569) and MARS endotypes (AUC: 0.477) in the GSE65682 datasets, the HLA classifier (AUC: 0.813) performed better in predicting mortality than age (AUC: 0.571) in the GSE95233 datasets, the HLA classifier (AUC: 0.752) performed better in predicting mortality than SRS endotypes (AUC: 0.570) and age (AUC: 0.675) in the E-MTAB-4421 datasets, the HLA classifier (AUC: 0.691) performed better in predicting mortality than age (AUC: 0.504) and SRS endotypes (AUC: 0.390) in the E-MTAB-4451 datasets, and the HLA classifier (AUC: 0.737) performed better in predicting mortality than age (AUC: 0.575),

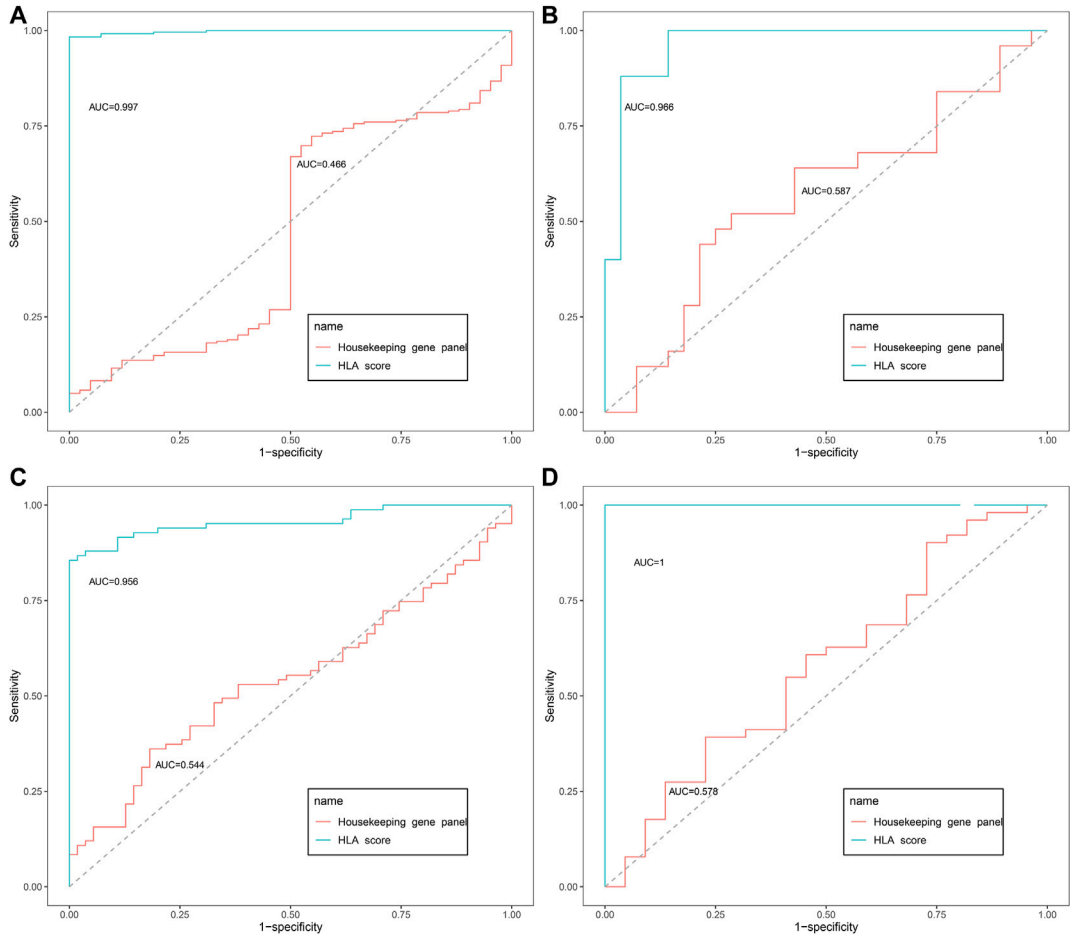


FIGURE 2 | The diagnostic efficacy of the HLA classifier in the discovery cohorts. **(A)** GSE65682 datasets. **(B)** GSE57065 datasets. **(C)** GSE69528 datasets. **(D)** GSE95233 datasets.

SRS endotypes (AUC: 0.534), and APACHE II score (AUC: 0.681) in the E-MTAB-7851 datasets (**Figure 4**).

Clinical Usefulness of the HLA Classifier

The E-MTAB-7581 dataset was collected from the VANISH randomized trial with patients randomized to receive either vasopressin or norepinephrine followed by placebo or hydrocortisone. In the HLA high-risk subgroup, the use of hydrocortisone (OR: 2.84, 95% CI 1.07–7.57, p = 0.037) was associated with increased risk of mortality. In the HLA low-risk subgroup, the use of hydrocortisone (OR: 1.63, 95% CI 0.21–12.71, p = 0.644) did not render significant alteration (**Table 3**). Notably, the DCA chart showed that the HLA classifier outperformed age, SRS, MARS, and APACHE II according to the net benefit of risk stratification using the model (y-axis) and the continuity of potential death threshold (x-axis) in the discovery and external validation cohorts (**Figure 5** and **Supplementary Figure S6**).

Immune Cell Infiltration Analysis

We analyzed the difference in composition of immune cells between the HLA subgroups in multiple transcriptome sets. The CIBERSORTx results demonstrated that compared with the HLA high-risk subgroup, activated memory CD4 T cells (p = 0.036), activated NK cells (p = 0.035), neutrophils (p = 0.001), and activated mast cells (p = 0.003) were more abundant in the HLA low-risk subgroup, while naive CD4 T cells (p = 0.003), regulatory T cells (Tregs) (p = 0.004), M0 macrophages (p = 0.001), and resting mast cells (p = 0.008) were more abundant in the high-risk subgroup than in the low-risk subgroup (**Figure 6A**) in GSE65682 datasets. In E-MTAB-4421 datasets, CD8 T cells (p < 0.001), resting memory CD4 T cells (p < 0.001), resting NK cells (p < 0.001), monocytes (p < 0.001), and activated dendritic cells (p = 0.001) were more abundant in the low-risk subgroup, but memory B cells (p < 0.001), naive CD4 T cells (p < 0.001), Tregs (p < 0.001), and M0 macrophages (p = 0.001) were more abundant in the high-risk subgroup than in the low-risk subgroup (**Figure 6B**). In E-MTAB-4451 datasets, the CIBERSORTx results uncovered that compared to the high-risk group, resting NK cells (p = 0.002) were more abundant in the low-risk subgroup, yet Tregs (p = 0.005) were more abundant in the high-risk subgroup than in the low-risk subgroup (**Figure 6C**). **Supplementary Figure S7**, **Supplementary Figure S8**, and **Supplementary Figure S9** display the distribution of 22 types of

TABLE 2 | Univariable and multivariable logistic regression analysis for prediction of survival in GEO and ArrayExpress databases.

Dataset	Factors	Subgroup	Univariable analysis		Multivariable analysis	
			OR (95%CI)	*p*	OR (95%CI)	*p*
GSE65682	—	—	—	—	—	—
	Age	—	1.02 (0.99–1.04)	0.114	NA	NA
	Sex	Female	1	—	—	—
		Male	1.42 (0.75–2.67)	0.284	NA	NA
	MARS	1–2	1	—	—	—
	—	3–4	0.93 (0.67–1.31)	0.688	NA	NA
	DM	No	1	—	—	—
	—	Yes	1.63 (0.71–3.74)	0.248	NA	NA
	HLA score	—	5.13 (2.03–15.39)	<0.001*	5.13 (2.03–15.39)	<0.001*
GSE63042	—	—	—	—	—	—
	HLA score	—	1.35 (4.22–43.18)	<0.001*	1.35 (4.22–43.18)	<0.001*
GSE54514	—	—	—	—	—	—
	HLA score	—	1.19 (1.06–1.45)	0.006*	1.19 (1.06–1.45)	0.006*
GSE95233	—	—	—	—	—	—
	Age	—	0.99 (0.95–1.03)	0.657	NA	NA
	Sex	Female	1	—	—	—
	—	Male	0.65 (0.35–1.18)	0.156	NA	NA
	HLA score	—	15.68 (3.04–80.93)	<0.001*	15.68 (3.04–80.93)	<0.001*
E-MTAB-4421	—	—	—	—	—	—
	Age	—	1.05 (1.02–1.07)	<0.001*	1.05 (1.03–1.08)	<0.001*
	Sex	Female	1	—	—	—
	—	Male	1.16 (0.64–2.10)	0.620	NA	NA
	SRS	1	—	—	—	—
	—	2	1.77 (0.98–3.20)	0.060	NA	NA
	HLA score	—	2.68 (1.44–4.97)	0.002*	2.93 (1.53–5.58)	0.001*
E-MTAB-4451	—	—	—	—	—	—
	Age	—	1.00 (0.98–1.03)	0.934	NA	NA
	Sex	Female	1	—	—	—
	—	Male	0.95 (0.40–2.28)	0.913	NA	NA
	SRS	1	—	—	—	—
	—	2	2.70 (1.18–6.19)	0.019	1.35 (0.50–3.70)	0.555
	HLA score	—	5.69 (2.08–15.58)	0.001*	5.69 (2.08–15.58)	0.001*
E-MTAB-7581	—	—	—	—	—	—
	Age	—	1.02 (0.99–1.04)	0.091	NA	NA
	Sex	Female	1	—	—	—
	—	Male	1.39 (0.71–2.72)	0.343	NA	NA
	SRS	1	—	—	—	—
	—	2	1.31 (0.68–2.55)	0.423	NA	NA
	APACHE II	—	1.08 (1.04–1.13)	0.001*	1.24 (1.06–1.19)	0.001*
	HLA score	—	11.35 (3.85–33.44)	<0.001*	17.74 (5.37–58.66)	<0.001*

*Abbreviations: OR, odds ratio; CI, confidence intervals; MARS, the Molecular Diagnosis and Risk Stratification of Sepsis; DM, diabetes mellitus, APACHE II, acute physiology and chronic health evaluation; SRS, sepsis response signature. NOTE: NA, not available. These variables were eliminated in the multivariate logistic regression model, so the HR, and p values were not available.****p** *< 0.05.*

immune cells in each sample for GSE65682 datasets, E-MTAB-4421 datasets, and E-MTAB-4451 datasets, respectively. In total, NK cells were significantly enriched in the HLA low-risk subgroup, whereas Tregs were more abundant in the HLA high-risk subgroup.

In addition, we adopted ssGSEA, another cell-type quantification method, to quantify the enrichment score of immune cell types. Compared to the CIBERSORTx results, the ssGSEA results revealed that significant infiltration of immune cells was concentrated in the HLA low-risk subgroup. In GSE65682 datasets, compared with the high-risk subgroup, activated B cells ($p < 0.001$), activated CD8 T cells ($p < 0.001$), NK cells ($p < 0.001$), activated dendritic cells ($p = 0.024$), T helper cells ($p < 0.001$), and infiltrating lymphocytes (IL) ($p < 0.001$) were more abundant in the low-risk subgroup, whereas Tregs ($p < 0.001$) and myeloid-derived suppressor cells (MDSCs) ($p = 0.002$) were more enriched in the high-risk subgroup than in the low-risk subgroup (**Figure 7A**). Similar results were observed in E-MTAB-4421 (**Figure 7B**) and E-MTAB-4451 datasets (**Figure 7C**), which indicated that patients in the HLA high-risk subgroup were characterized by immunosuppression.

Correlation Between HLA Classifier/Genes and Immune Cells

We further explored whether our HLA classifier/genes were related to immune cell infiltration in sepsis via Spearman correlation analyses in multiple gene expression profiles. In the GSE65682 dataset, the HLA score was significantly

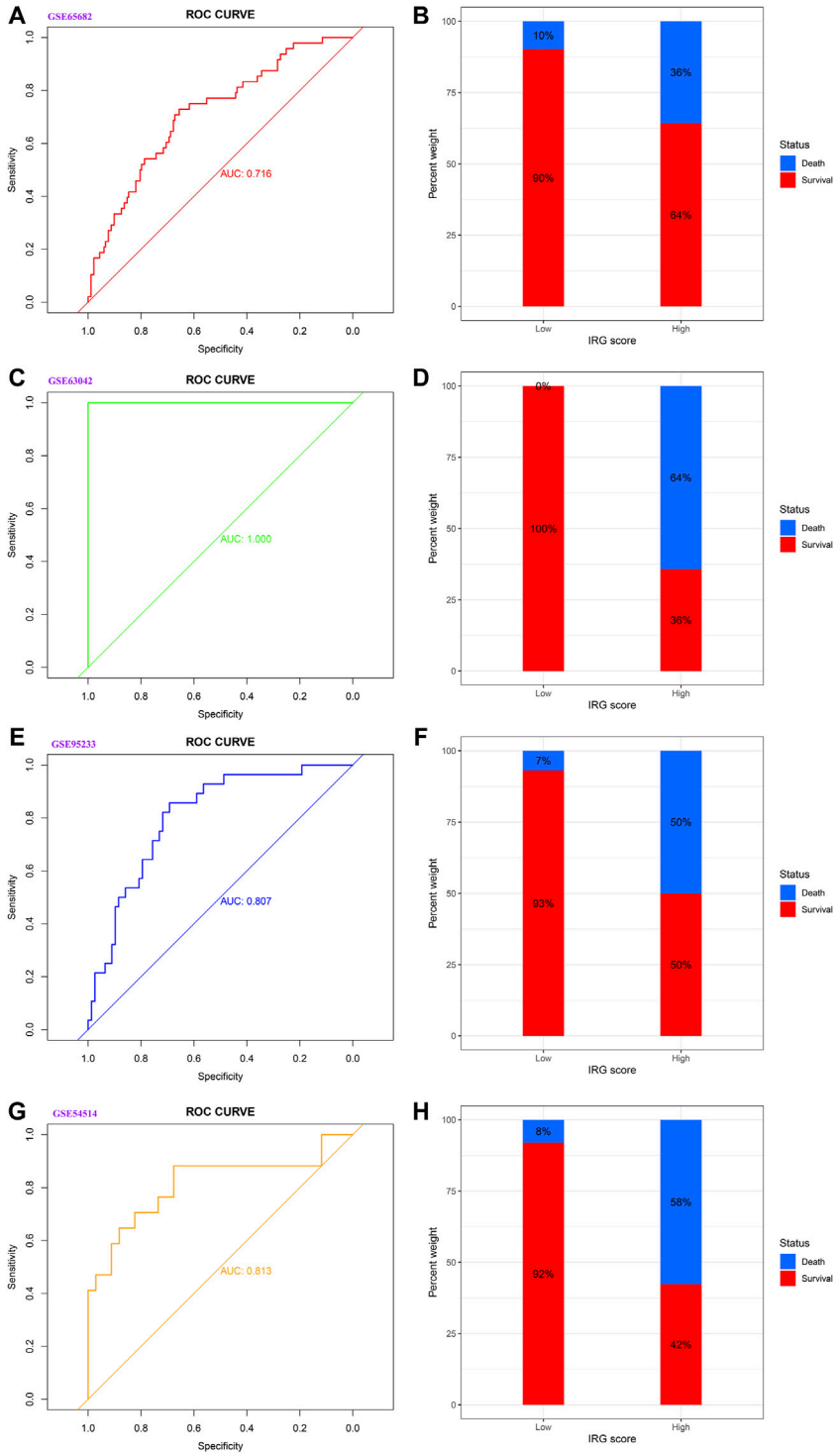


FIGURE 3 | The prognostic capacity of the HLA classifier and the distribution of mortality rate in different HLA subgroups in the discovery cohorts. **(A,B)** GSE65682 datasets. **(C,D)** GSE63042 datasets. **(E,F)** GSE95233 datasets. **(G,H)** GSE54514 datasets.

negatively correlated with B cells ($p < 0.001$), NK cells ($p < 0.001$), activated dendritic cells ($p<0.01$), ILs ($p < 0.001$), and T helper cells ($p < 0.001$), whereas the HLA score was significantly positively correlated with Tregs ($p < 0.001$) and MDSCs ($p < 0.05$) (**Figure 8A** and **Supplementary Table S5**). Likewise, in the E-MTAB-4421 dataset, the HLA score was significantly negatively correlated with B cells ($p < 0.001$), NK cells ($p < 0.001$), activated dendritic cells ($p <0.001$), ILs ($p <$

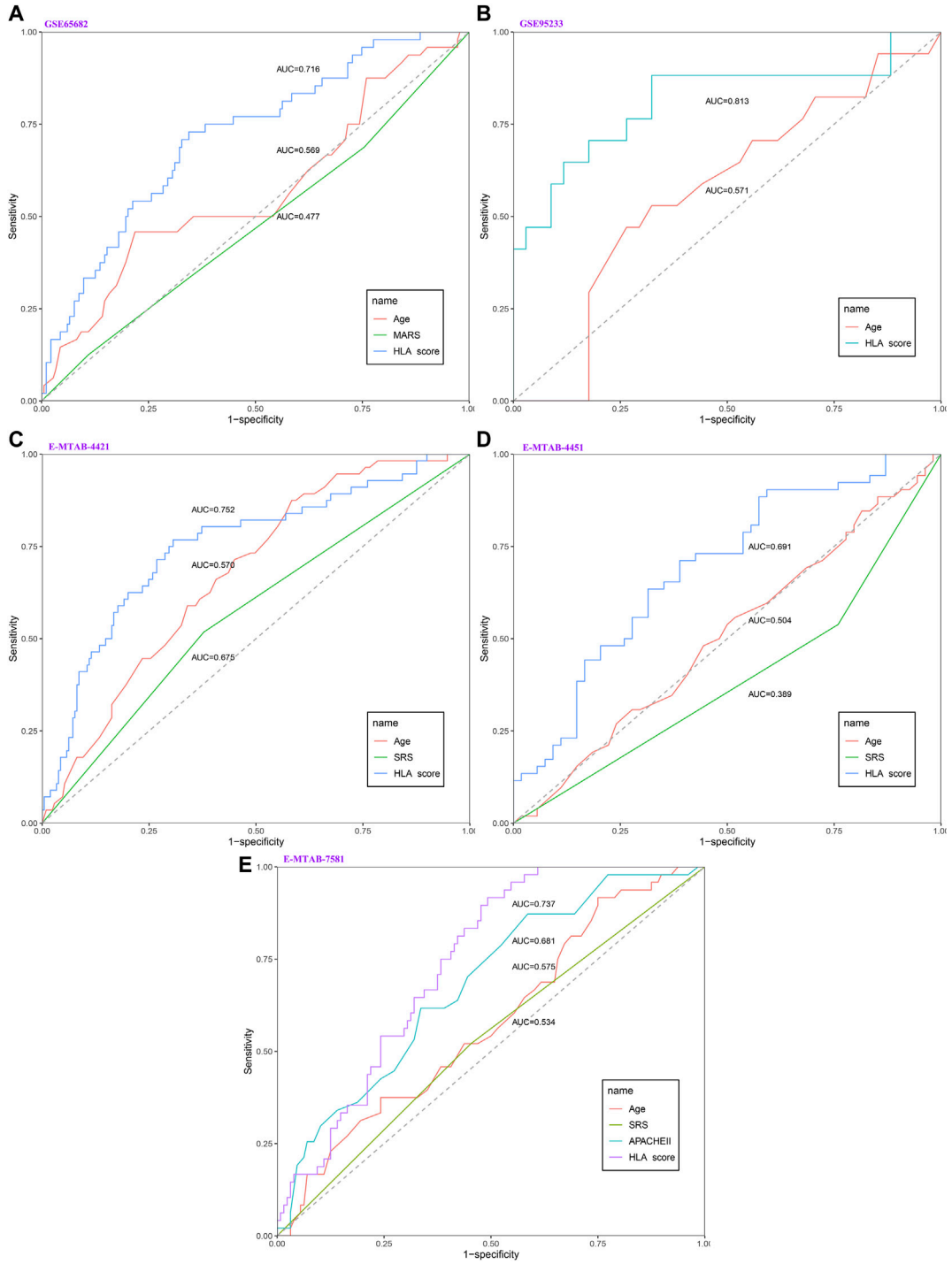


FIGURE 4 | The evaluation of the performance of the HLA classifier compared against age, SRS, MARS, and APACHE II in the discovery and external validation cohorts. **(A)** GSE65682 datasets. **(B)** GSE95233 datasets. **(C)** E-MTAB-4421 datasets. **(D)** E-MTAB-4451 datasets. **(E)** E-MTAB-7581 datasets.

0.001), and T helper cells ($p < 0.001$), whereas the HLA score was significantly positively correlated with Tregs ($p < 0.001$) and MDSCs ($p < 0.001$) (**Figure 8B** and **Supplementary Table S6**). Similarly, in the E-MTAB-4451 dataset, the HLA score was significantly negatively correlated with B cells ($p < 0.001$), NK cells ($p < 0.01$), activated dendritic cells ($p < 0.001$), ILs ($p < 0.001$), and T helper cells ($p < 0.001$), whereas the HLA score was significantly positively correlated with Tregs ($p < 0.01$) and MDSCs ($p < 0.001$) (**Figure 8C** and **Supplementary Table S7**). In addition, **Figure 8** shows that B2M, HLA-DQA1,

TABLE 3 | Comparisons of the predictive value of the HLA classifier versus disease severity and SRS.

Models	Univariable analysis	
	OR (95% CI)	*p*
Use of hydrocortisone in the HLA low-risk subgroup	1.63 (0.21–12.71)	0.644
Use of hydrocortisone in the HLA high-risk subgroup	2.84 (1.07–7.57)	0.037*
Use of hydrocortisone in SRS 2	3.76 (1.41–10.04)	0.008*
Use of hydrocortisone in SRS 1	1.25 (0.47–3.36)	0.658
Use of hydrocortisone by APACHE II	0.94 (0.86–1.03)	0.210
Use of vasopressin in SRS 2	0.69 (0.18–2.62)	0.583
Use of vasopressin in SRS 1	1.50 (0.40–3.89)	0.403
Use of vasopressin in the HLA low-risk subgroup	2.91 (0.29–29.45)	0.366
Use of vasopressin in the HLA high-risk subgroup	1.26 (0.58–2.76)	0.559
Use of vasopressin by APACHE II	0.96 (0.87–1.06)	0.427

*The logistic regression models integrating interactions between treatment allocation and SRS, HLA, classifier, or APACHE II, were built in the E-MTAB-7581, dataset. A total of six logistic regression models were built by using mortality as the response variable and respective predictors and interactions were: hydrocortisone & class, hydrocortisone and SRS, hydrocortisone and APACHE II, vasopressin and class, vasopressin and SRS, and vasopressin and APACHE II. A significant (***p*** <0.05) interaction indicated that the classification method was of predictive value because it identified that a subgroup of patients responded differently to treatment. SRS, classification was used as previously reported. Abbreviations: OR, odds ratio; CI, confidence intervals; MARS, the Molecular Diagnosis and Risk Stratification of Sepsis; DM, diabetes mellitus, APACHE II, acute physiology and chronic health evaluation; SRS, sepsis response signature.*

HLA-DPA1, TAP1, and TAP2 were significantly associated with the infiltration of immune cells, HLA-DPA1 and HLA-DQA1 in particular. Nevertheless, the housekeeping gene panel is not correlated with immune cell infiltration (**Figure 8**).

Immune and Molecular Function Between the HLA Subgroups by GSVA and ssGSEA

To screen biological differences between the HLA subgroups, GSVA was conducted to determine the gene sets enriched in different HLA subgroups. In GSE65682 sets (**Supplementary Figure S10A**), the results showed that the NOD-like receptor signaling pathway, toll-like receptor signaling pathway, and complement and coagulation cascades were enriched in the HLA high-risk group, yet alanine aspartate and glutamate metabolism and glyoxylate and dicarboxylate metabolism were mainly involved in the HLA low-risk group. In GSE95233 sets (**Supplementary Figure S10B**), primary immunodeficiency, the PPAR signaling pathway, and complement and coagulation cascades were mainly enriched in the HLA high-risk group, but antigen processing and presentation and aminoacyl tRNA biosynthesis were involved in the HLA low-risk subgroup. In E-MTAB-4421 sets (**Supplementary Figure S10C**), spliceosome, valine leucine, and isoleucine degradation and RNA degradation were enriched in the HLA high-risk subgroup, while starch and sucrose metabolism, leukocyte transendothelial migration, and neuroactive ligand receptor interaction were mainly involved in the HLA low-risk subgroup. To sum up, complement and coagulation cascades may play an important role in the initiation and progression of sepsis.

In addition, ssGSEA was utilized to investigate the given immune-related pathway in sepsis. As a result, in GSE65682 datasets, all of the significantly different immune-related gene sets were enriched in the HLA low-risk subgroup, such as cytokine-cytokine receptor (CCR) interaction, cytolytic activity, human leukocyte antigen (HLA), inflammation–promoting, MHC class I, antigen processing machinery, antigen-presenting cell (APC) costimulation, parainflammation, the NF–kappa B signaling pathway, and the JAK–STAT signaling pathway (**Figure 9A**). Analogously, in GSE63042 datasets, the gene sets of the HLA low-risk group were enriched in CCR, HLA, inflammation–promoting, antigen processing machinery, APC coinhibition, APC costimulation, parainflammation, IL6 JAK–STAT3 signaling, the NF–kappa B signaling pathway, and the JAK–STAT signaling pathway (**Figure 9B**). Homoplastically, in E-MTAB-4421 datasets, the ssGSEA results showed that CCR, cytolytic activity, HLA, inflammation-promoting, and APC costimulation were involved in the HLA low-risk group (**Figure 9C**). Similarly, in E-MTAB-4451 datasets, the ssGSEA results demonstrated that CCR, cytolytic activity, HLA, inflammation-promoting, parainflammation, and APC costimulation were mainly enriched in the HLA low-risk group (**Figure 9D**). In short, the HLA high-risk group, compared with the HLA low-risk group, was characterized by immunosuppression in which many pivotal immune pathways were suppressed such as CCR, HLA, inflammation-promoting, and APC costimulation.

Correlation Between HLA Classifier/Genes and Pivotal Molecular Pathways

We further tested whether our HLA classifier/genes were related to molecular pathways in sepsis via Spearman correlation analyses in multiple transcriptome sets. Encouragingly, the HLA score was significantly negatively correlated with HLA (p <0.001), APC costimulation (p < 0.001), parainflammation (p < 0.001), antigen processing machinery (p < 0.001), and CCR (p < 0.001) in GSE65682 sets; HLA (p < 0.001), APC costimulation (p < 0.001), inflammation-promoting (p < 0.001), antigen processing machinery (p < 0.001), and CCR (p < 0.001) in GSE63042 sets; HLA (p < 0.001), APC costimulation (p < 0.001), inflammation-promoting (p < 0.001), MHC class I (p < 0.001), and CCR (p <

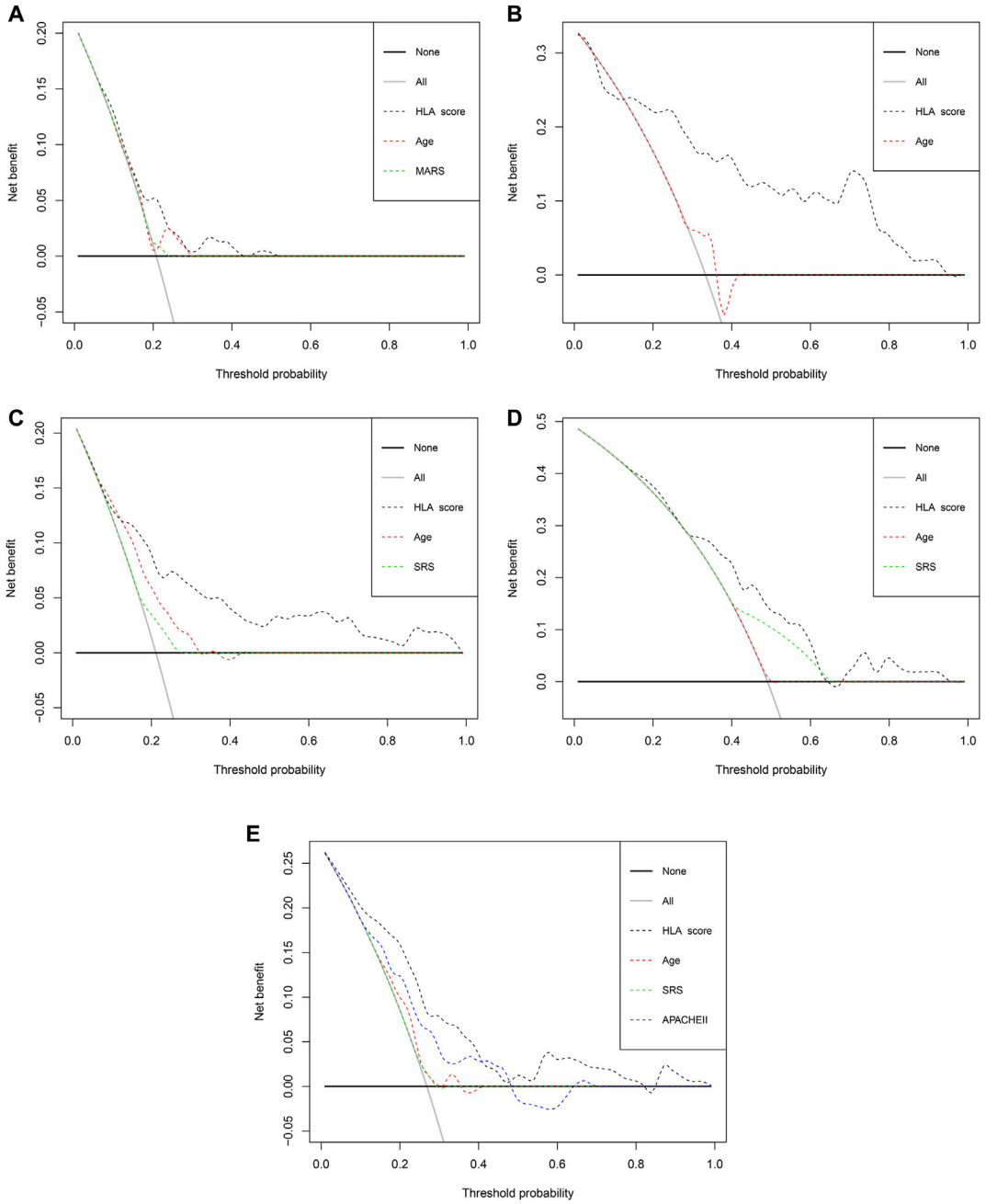


FIGURE 5 | The evaluation of the clinical usefulness of the HLA classifier compared with age, SRS, MARS, and APACHE II in the discovery and external validation cohorts. **(A)** GSE65682 datasets. **(B)** GSE95233 datasets. **(C)** E-MTAB-4421 datasets. **(D)** E-MTAB-4451 datasets. **(E)** E-MTAB-7581 datasets.

0.001) in E-MTAB-4421 sets; and HLA ($p < 0.001$), APC costimulation ($p < 0.05$), inflammation-promoting ($p < 0.001$), antigen processing machinery ($p < 0.01$), and CCR ($p < 0.05$) in E-MTAB-4451 sets (**Figure 10**). In addition, **Figure 10** shows that B2M, HLA-DQA1, HLA-DPA1, TAP1, and TAP2 were significantly associated with the enrichment score of immune-related pathways, TAP1 and TAP2 in particular. Nevertheless, the housekeeping gene panel is not correlated with immune-related pathways (**Figure 10**).

Analyses of the Cytokines

To analyze the clinically detectable inflammatory cytokines involved in sepsis, we applied the Wilcoxon test to compare the expression levels of cytokines in different HLA endotypes. As a result, in GSE65682 sets, the expression levels of CCL5, IL1B, and IL15 were significantly higher in the HLA low-risk group, but the expression levels of IL10 were significantly downregulated in the HLA low-risk group (**Figure 11A**). Analogously, in GSE63042 sets, the expression levels of IL1B, TNF and VEGFA were

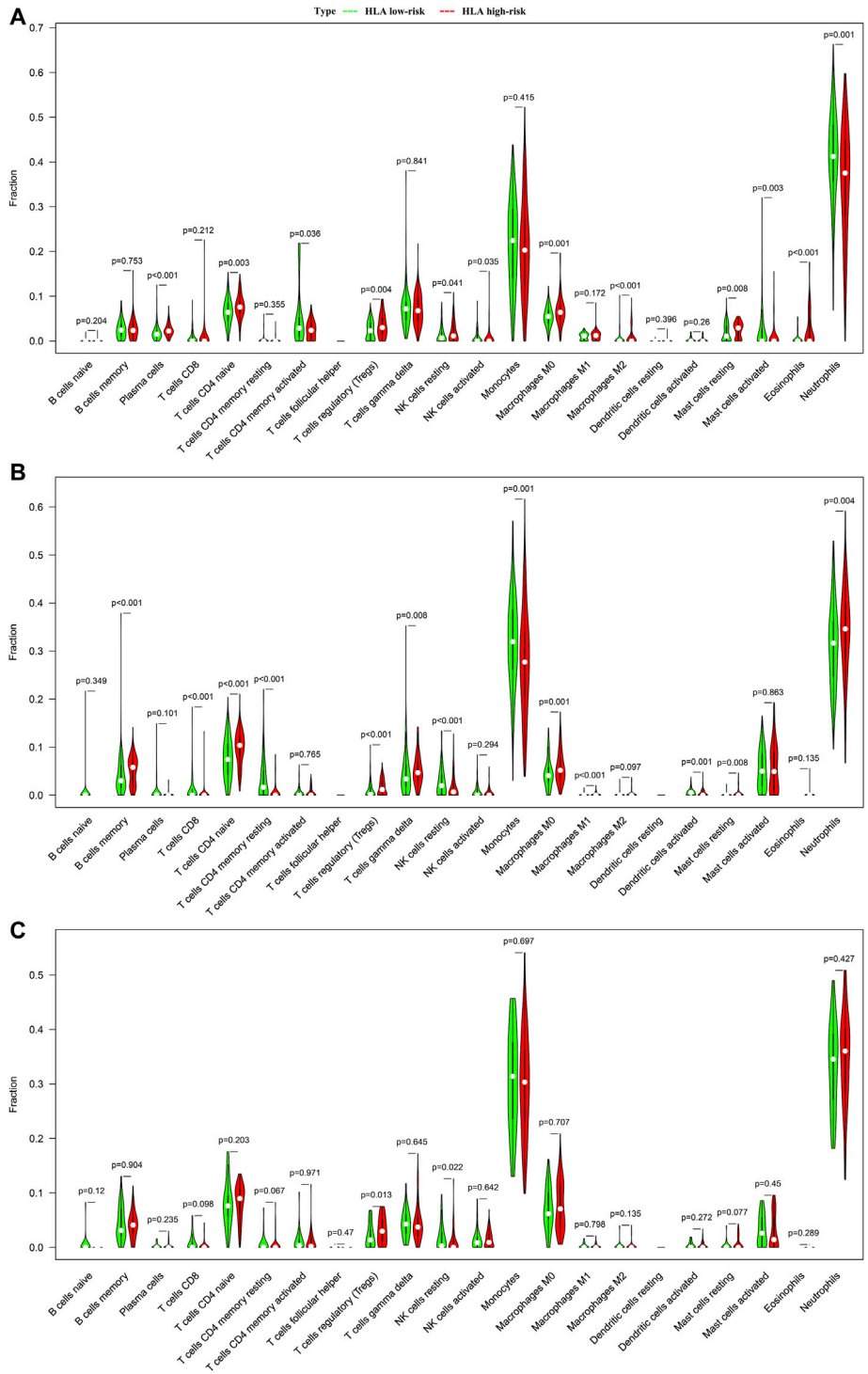


FIGURE 6 | Comparison of infiltrating immune cells between different HLA subgroups based on CIBERSORTx in multiple transcriptome datasets. **(A)** GSE65682 datasets. **(B)** E-MTAB-4421 datasets. **(C)** E-MTAB-4451 datasets. Green indicates HLA low-risk, while red indicates HLA high-risk.

significantly higher in the HLA low-risk group, but the expression levels of IL10 were significantly lower in the HLA low-risk group (**Figure 11B**). Similarly, in E-MTAB-4421 sets, the expression levels of CCL5, CXCL10, IFNG, and PDGFRB were significantly higher in the HLA low-risk group, and the expression levels of TNF exhibited a trend toward a higher expression in the HLA low-risk group, but the expression levels of CCL11, IL10, and IL1RN were significantly lower in the HLA low-risk group

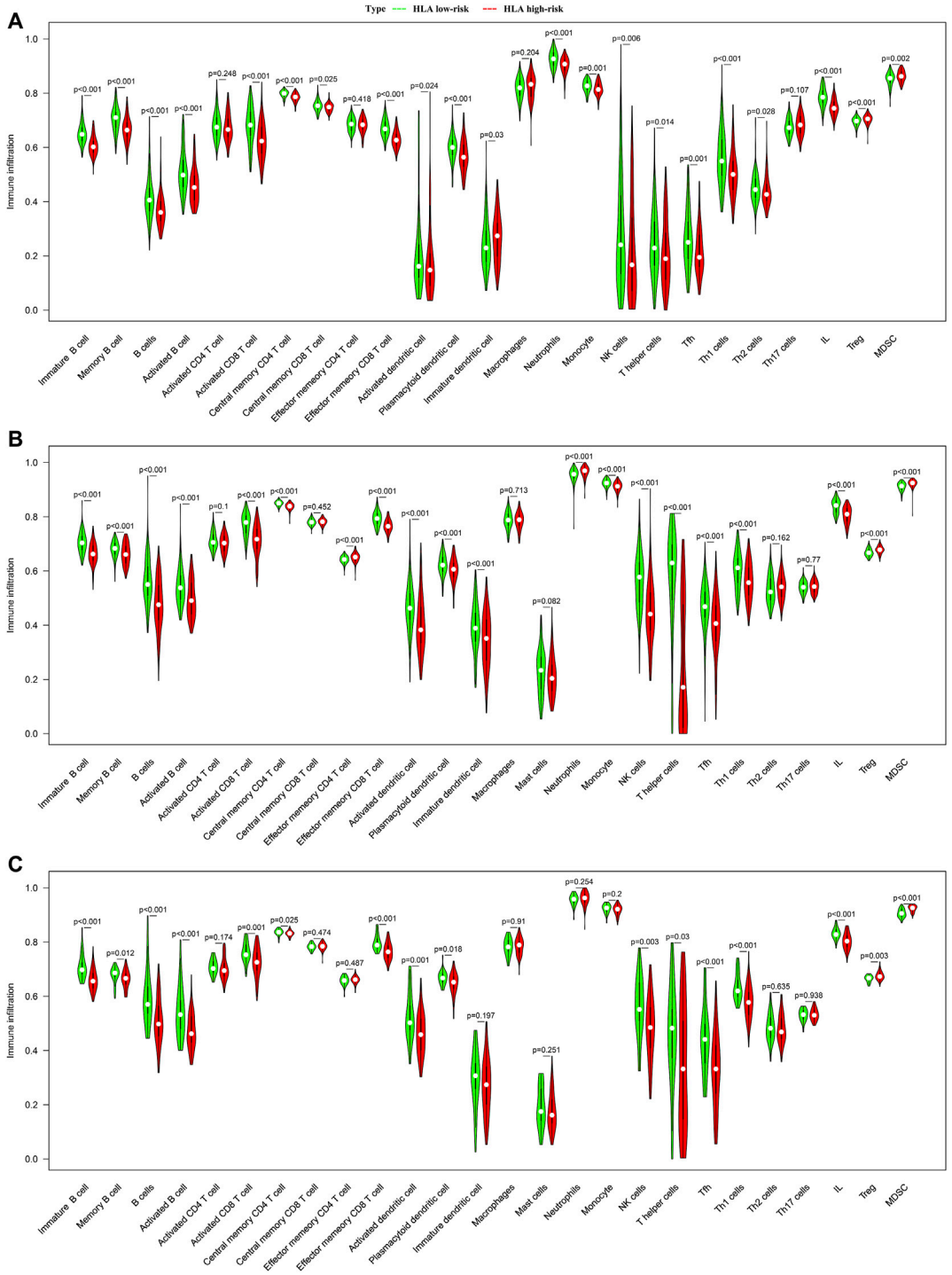


FIGURE 7 | Comparison of infiltrating immune cells between different HLA subgroups based on ssGSEA in multiple transcriptome datasets. **(A)** GSE65682 datasets. **(B)** E-MTAB-4421 datasets. **(C)** E-MTAB-4451 datasets. Green indicates HLA low-risk, while red indicates HLA high-risk.

(**Figure 11C**). In summary, pro-inflammatory cytokines were upregulated in the HLA low-risk subgroup and anti-inflammatory cytokines were upregulated in the HLA high-risk subgroup.

In addition, we further explored whether our HLA classifier was associated with the ratio of IL10/TNF in sepsis. As a results, the HLA score was significantly positively correlated with IL10/TNF ($R = 0.36$, $p < 0.001$) in GSE65682 sets, positively correlated

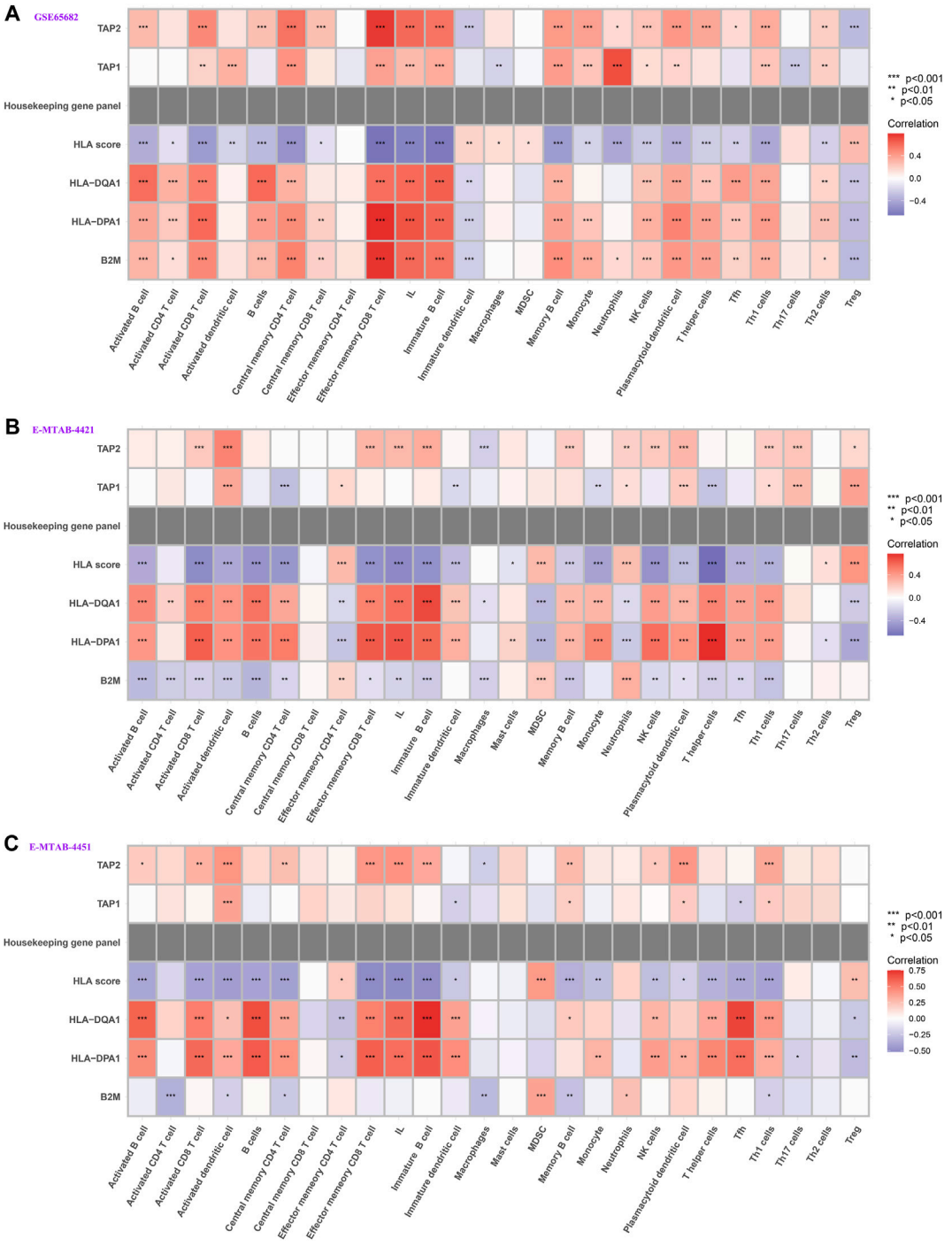


FIGURE 8 | Correlation between HLA classifier/genes and immune cells in multiple transcriptome datasets. **(A)** GSE65682 datasets. **(B)** E-MTAB-4421 datasets. **(C)** E-MTAB-4451 datasets.

(R = 0.13, p = 0.037) in GSE63042 sets, and positively correlated (R = 0.3, p = 0.0018) in E-MTAB-4421 sets (**Figures 12A–C**).

qRT-PCR

To further validate the expression of the five HLA genes, we performed qRT-PCR in 50 clinical specimens. Compared with healthy controls, B2M was significantly upregulated in sepsis samples (**Figure 12D**) yet the expression level of HLA-DQA1, HLA-DPA1, TAP1, and TAP2 was significantly lower in sepsis specimens. Collectively, the results of qRT-PCR were in accordance with the results of bioinformatics analyses derived from the GEO datasets.

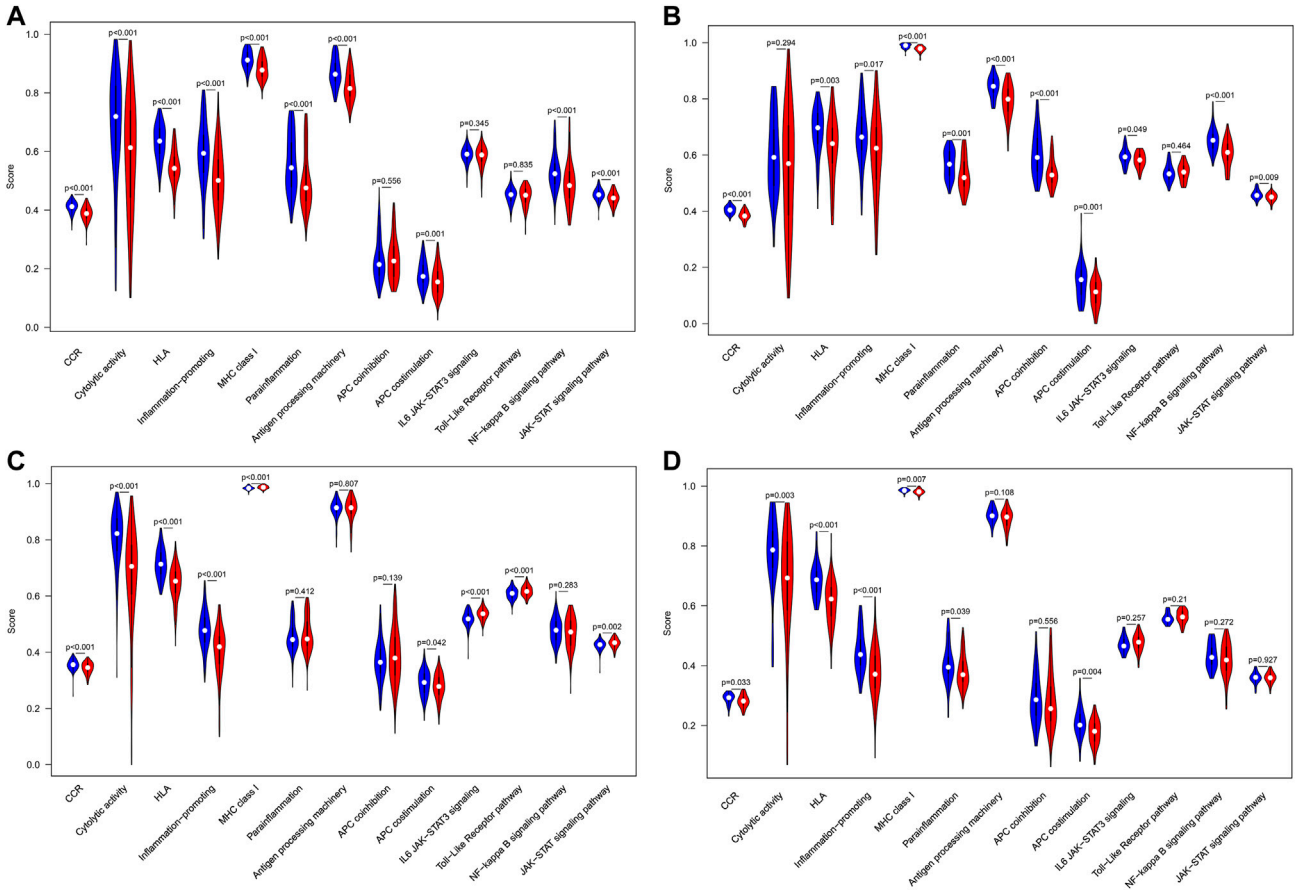


FIGURE 9 | Comparison of immune-related pathways between different HLA subgroups in the discovery cohorts. **(A)** GSE65682 datasets. **(B)** GSE63042 datasets. **(C)** E-MTAB-4421 datasets. **(D)** E-MTAB-4451 datasets. Blue indicates HLA low-risk, while red indicates HLA high-risk.

DISCUSSION

After analyzing multiple gene expression profiling, according to modified Lasso penalized regression and RF, five HLA genes (B2M, HLA-DQA1, HLA-DPA1, TAP1, and TAP2) were identified as hub genes, which were used to construct a prediction model, namely the HLA classifier. In the discovery cohort, the HLA classifier exhibited superior diagnostic efficacy (AUC = 0.997) and performed better in predicting mortality (AUC = 0.716) than clinical characteristics or MARS/SRS endotypes. Encouragingly, similar results were observed in the ArrayExpress databases. In the E-MTAB-7581 dataset, the use of hydrocortisone in the HLA high-risk subgroup (OR: 2.84, 95% CI 1.07–7.57, p = 0.037) was associated with increased risk of mortality. Immune infiltration analysis by CIBERSORTx showed that NK cells were significantly enriched in the HLA low-risk subgroup, while Tregs were more abundant in the HLA high-risk subgroup. Intriguingly, ssGSEA also revealed that B cells, activated dendritic cells, NK cells, T helper cells, and ILs were significantly enriched in the HLA low-risk subgroup, while Tregs and MDSCs were more abundant in the HLA high-risk subgroup. The HLA score was significantly negatively correlated with the infiltration score of B cells, activated dendritic cells, NK cells, T helper cells, and ILs, yet was significantly positively correlated with the infiltration score of Tregs and MDSCs. Additionally, molecular pathways determined via the ssGSEA algorithm uncovered that CCR, HLA, and APC costimulation was significantly enriched in the HLA low-risk subgroup, enrichment scores of which were significantly negatively correlated with HLA score. Finally, the expression levels of several cytokines (IL-10, IFNG, TNF) were significantly different between the HLA phenotypes, and the ratio of IL-10/TNF was significantly positively correlated with HLA score. Results of qRT-PCR validated the higher expression level of B2M as well as lower expression level of HLA-DQA1, HLA-DPA1, TAP1, and TAP2 in sepsis samples compared to control samples.

To the best of our knowledge, this is the first comprehensive study to explore HLA gene sets based on a multiple transcriptome expression profiles in all-cause sepsis, leading to the discovery of novel biomarkers to develop a diagnostic and prognostic model, thus elucidating the model and immune system (immune cell infiltration, immune-related pathways, and cytokines) to find its additional clinical implications.

At present, no single biomarker can be efficient in diagnosing sepsis, prognosis, and monitoring disease with especially high

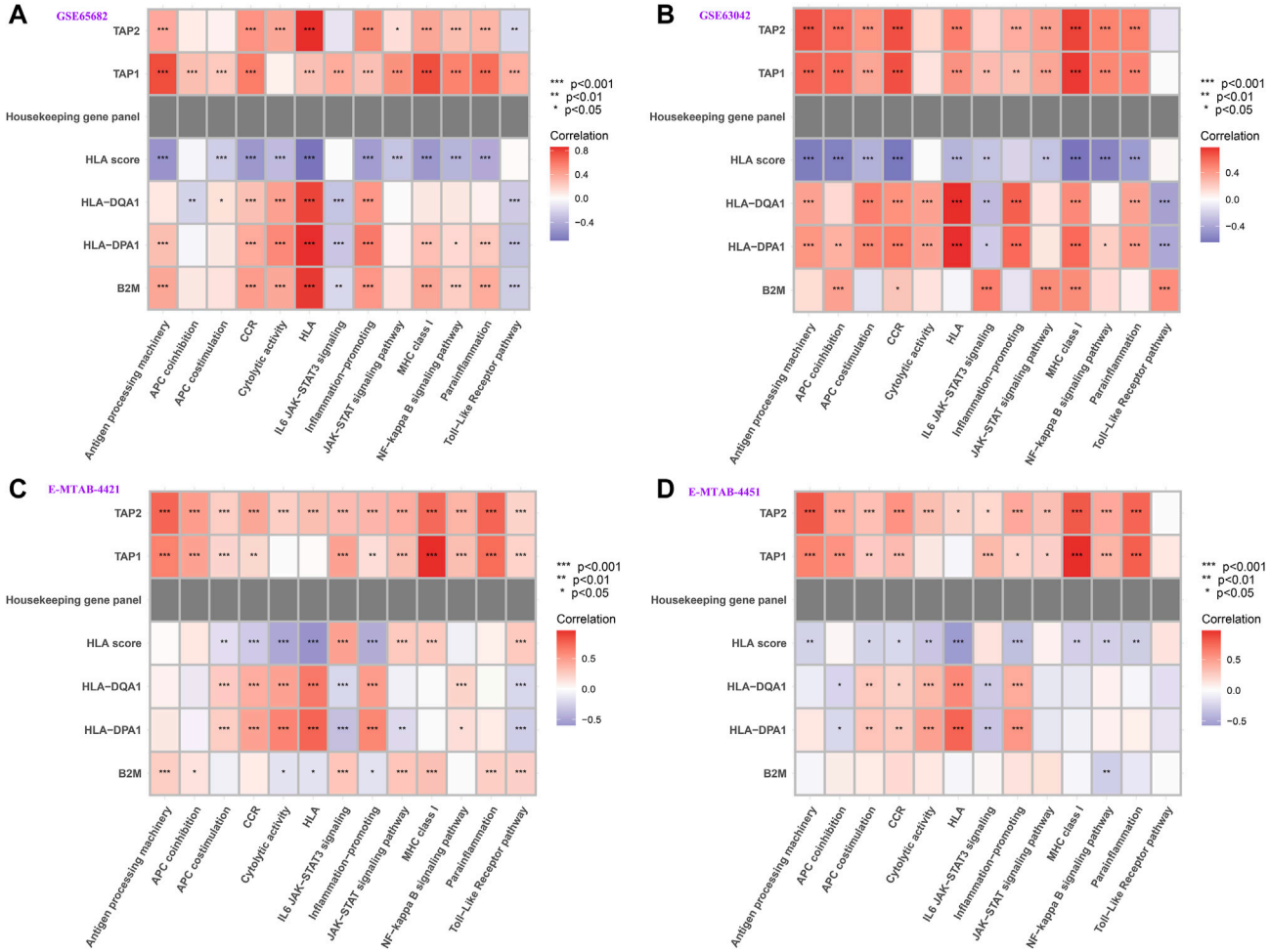


FIGURE 10 | Correlation between HLA classifier/genes and immune-related pathways in multiple transcriptome datasets. **(A)** GSE65682 datasets. **(B)** GSE63042 datasets. **(C)** E-MTAB-4421 datasets. **(D)** E-MTAB-4451 datasets.

performance uniformly according to the variety of factors and processes involved in sepsis (Jensen and Bouadma, 2016). This is most likely due to heterogeneity in the adult host response to infection and fails to capture important pathophysiological alterations, thus cannot uncover underlying mechanisms. HLA gene sets, as promising novel biomarkers, may offer important predictive and prognostic information. Machine learning methods, which can decrease diagnostic uncertainties and analyze the heterogeneity in transcriptome data (Baniasadi et al., 2021), including RF based on minimum error regression trees and modified Lasso coupled with adequate validation metrics, were applied to identify reliable feature variables. Based on RF and Lasso, five HLA genes (B2M, HLA-DQA1, HLA-DPA1, TAP1, and TAP2) were identified as hub genes, which were combined to construct an HLA classifier. As to diagnostic ability, the AUC of the HLA classifier was more than 0.95 in multiple transcriptome sets, which demonstrated that the HLA classifier can efficiently discriminate sepsis from the control samples. As for prognostic capacity, the HLA classifier was an independent predictor of unfavorable survival outcome, regardless of other clinical characteristics, in multiple transcriptome datasets. Importantly, the performance of the HLA classifier in predicting mortality outcomes was superior to clinical features or MARS/SRS endotypes. In total, the model, HLA classifier, could be a robust tool to diagnose sepsis earlier and to identify patients at risk of a poor or even fatal outcome.

Up to now, prognostic biomarkers/models have mainly been utilized for overall prognosis, which is not enough. Added information should include how to stratify patients to guide treatment. Interestingly, our results found that though the HLA classifier could not modify the effect of norepinephrine versus vasopressin, the HLA high-risk subgroup exhibited a significantly higher mortality outcome when assigned to the hydrocortisone group, consistent with the GAinS study where the use of hydrocortisone in SRS1, which represents an immunosuppressed phenotype including features of downregulation of HLA class II, endotoxin tolerance, and T-cell exhaustion, was associated with increased risk of mortality (Davenport et al., 2016). The probable explanation is that the HLA low-risk subgroup was relatively

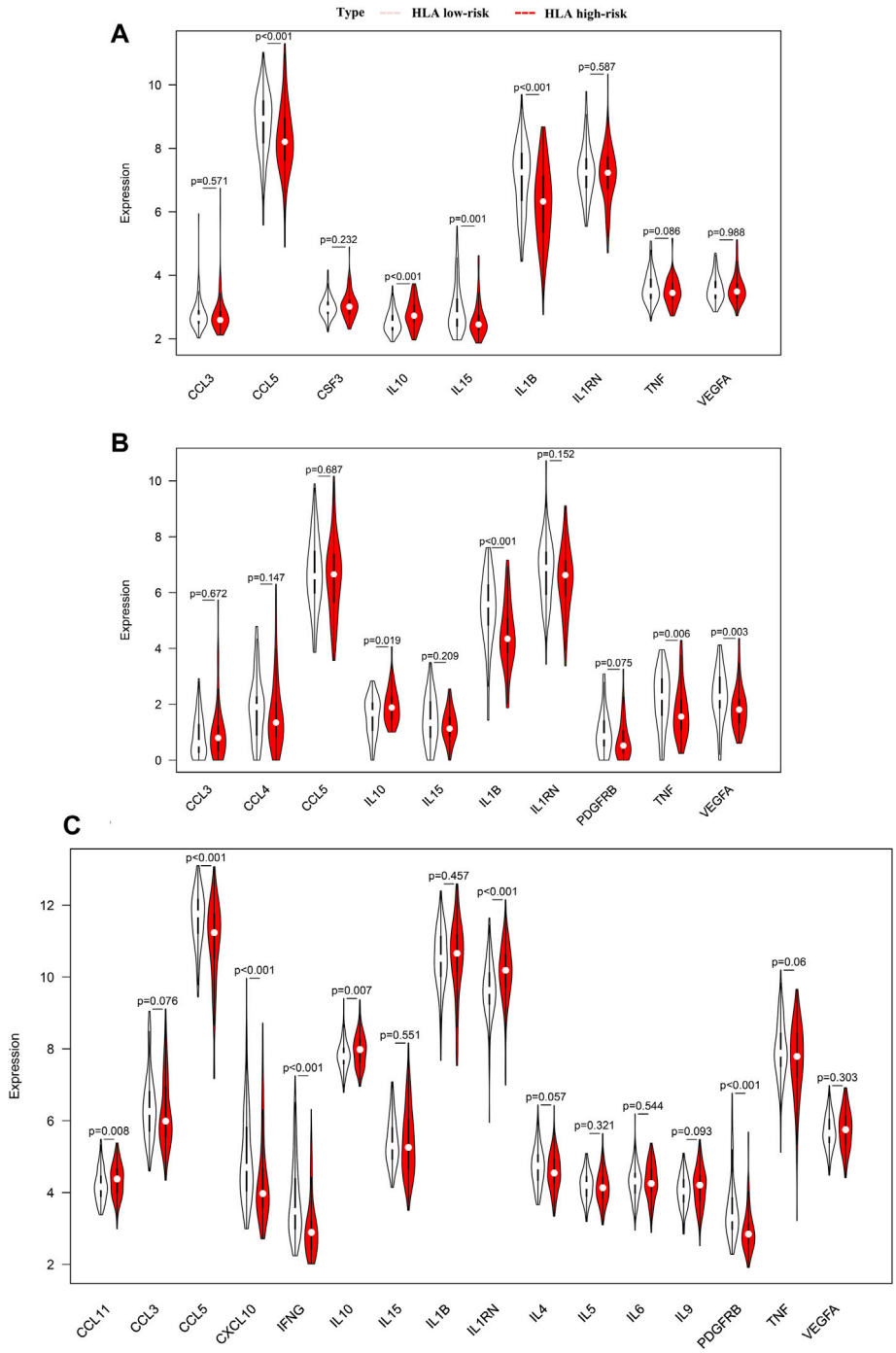


FIGURE 11 | Comparison of the expression level of cytokines between different HLA subgroups in multiple transcriptome datasets. **(A)** GSE65682 datasets. **(B)** GSE63042 datasets. **(C)** E-MTAB-4421 datasets. Pink indicates HLA low-risk, while red indicates HLA high-risk.

immunocompetent with a lower mortality rate, and the HLA high-risk subgroup was relatively immunocompromised with a higher mortality rate. The use of hydrocortisone suppresses the immune system (Steinhagen et al., 2020), which aggravates the immunosuppression status of the HLA high-risk subgroup, thereby increasing the mortality rate. The HLA high-risk subgroup may not be suitable for the application of hydrocortisone. Additionally, DCA results indicated that survival-associated treatment decisions for sepsis patients based on the HLA classifier had a net benefit compared to treatment decisions based on other clinical features or MARS/SRS endotypes, or treatment for all patients or none. To sum up,

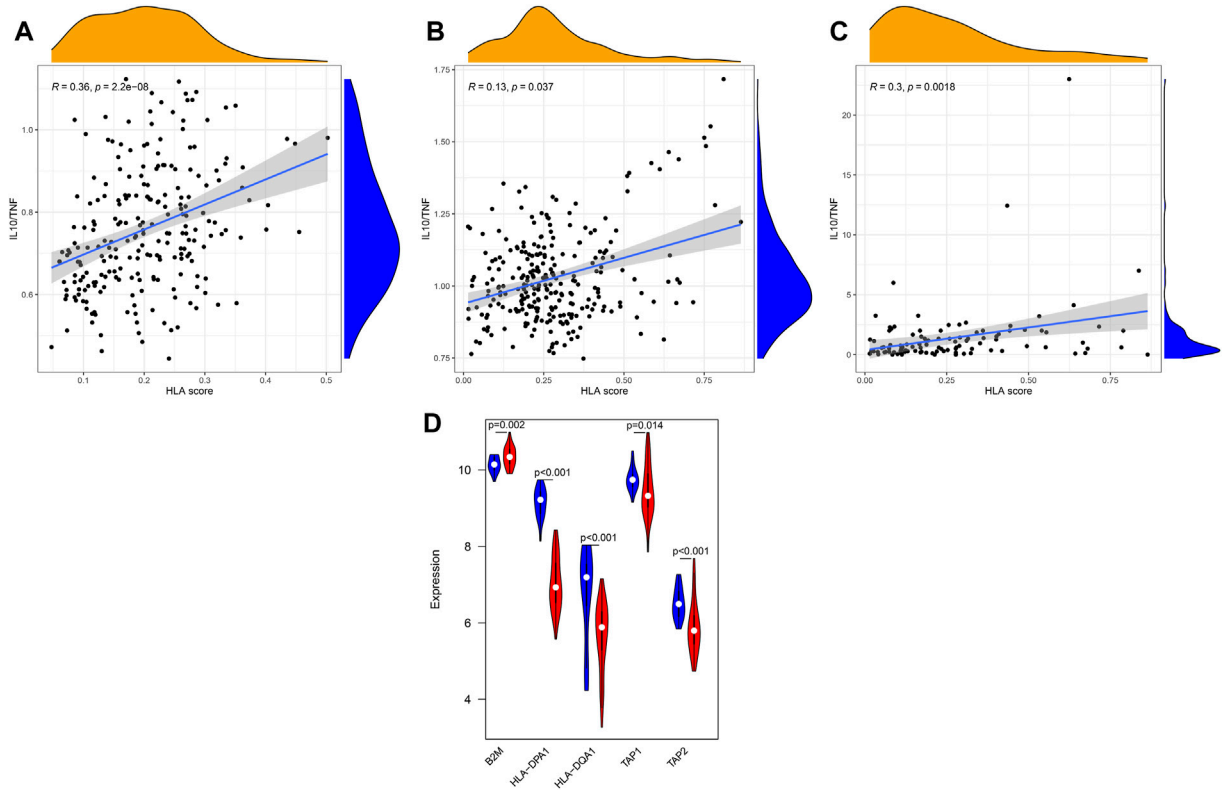


FIGURE 12 | Correlation between the HLA classifier and the ratio of IL10/TNF in multiple transcriptome datasets. **(A)** GSE65682 datasets. **(B)** GSE63042 datasets. **(C)** E-MTAB-4421 datasets. **(D)** Comparison of gene expression levels of five HLA genes between controls and sepsis samples *via* qRT-PCR. Red indicates sepsis, while blue indicates normal.

the current HLA classifier could be useful for clinicians to tailor survival-related treatment decisions.

Excessive immune activation and concurrent immunosuppression are central to the pathophysiology of sepsis. Immunosuppression results in a profound dysfunction in innate and adaptive immune responses, which mainly manifests as the depletion and exhaustion of lymphocytes, increased apoptosis of immune cells, the expansion of Treg cells and MDSCs, downregulation of activating cell-surface molecules (HLA-DR), and inhibitory proinflammatory cytokine release (Hotchkiss et al., 2013). It is becoming increasingly clear that most sepsis patients are not succumbing to an overwhelming pro-inflammatory response early on, but rather to immunoparalysis-related complications that occur later in the disease trajectory (Cui et al., 2019). The severe suppression status of the immune system hampers the patient from clearing the primary infection and increases susceptibility toward secondary and opportunistic infections, thereby leading to many adverse clinical consequences.

Currently, mHLA-DR is a reliable biomarker for evaluating immunosuppression and is widely utilized to guide immunomodulation therapies. Unfortunately, innumerable clinical trials of promising immunostimulation therapies have failed to achieve the desired effect and the consensus is that heterogeneity, especially in individual immune statuses, is responsible for these dismal failures. Due to single biomarkers with limited statistical power, multiple molecular signatures appear to provide better predictive information. Surprisingly, the HLA classifier is closely associated with the immunesuppressive state from multiple perspectives, including infiltrating immune cells, immune-related pathways, and cytokines level, which may act as an effective indicator of immunological paralysis.

One hallmark is apoptosis of B cells and dendritic cells and the depletion and exhaustion of T lymphocytes during sepsis-induced immunoparalysis resulting in an acquired immune deficiency syndrome that is associated with poor outcomes (Hotchkiss et al., 2013). Similarly, deficiency of T helper cells (Th1, Th2, and Th17 cells) proves detrimental to sepsis patients by promoting immunoparalysis, which is associated with increased mortality (Wu et al., 2013). Analogously, the reduced NK cell number and dysfunction may impair the host's defense against pathogens and make them more vulnerable to nosocomial infection, which participates in sepsis-induced immunosuppression (Kjaergaard et al., 2015). In our study, B cells, activated dendritic cells, ILs, T helper cells, and NK cells were more abundant in HLA low-risk phenotypes than in HLA high-risk phenotypes, and were significantly negatively correlated with the HLA classifier, which is in accordance with the feature of immunosuppression. Conversely, Treg cells that are upregulated in the immunoparalysis stage of sepsis, maintain self-tolerance via inhibiting/suppressing neutrophils, monocytes, and effector T cells, which are associated with clinical worsening and

mortality (Kumar, 2018). Likewise, MDSCs, a heterogeneous population of inducible immature myeloid cells with immunosuppressive properties (such as inducing the expansion of Treg cells and suppressing T-cell responses), are expanded during sepsis and serve as one of the contributing factors for sepsis-associated mortality (Schrijver et al., 2019). In our research, Treg cells and MDSCs were significantly more enriched in HLA high-risk endotypes than in HLA low-risk endotypes, and were significantly positively correlated with the HLA classifier, which is in accordance with the characteristic of immunoparalysis. In total, the HLA classifier is negatively associated with activated immune cells defending against infectious, while is positively associated with immunosuppression cells.

Intriguingly, from multiple transcriptome profiles, all of the different immune-related gene sets were significantly enriched in HLA low-risk phenotypes, such as CCR, cytolytic activity, HLA, inflammation-promoting, parainflammation, MHC class I, antigen processing machinery, and APC costimulation, particularly CCR, inflammation-promoting, APC costimulation, and HLA. That is to say, HLA high-risk endotypes were characterized by immunosuppression where numerous activated immune pathways were inhibited compared to HLA low-risk endotypes. In addition, the HLA classifier was significantly negatively associated with CCR, inflammation-promoting, APC costimulation, and HLA, which hints that the HLA classifier can serve as a surrogate marker of sepsis-induced immunosuppression.

Cytokines are one of the key causes underlying sepsis-related immunosuppression and produced by immune cells. During sepsis, a maladjusted and excessive release of pro-inflammatory and anti-inflammatory cytokines will result in a cytokine storm in the early stage of sepsis. However, in the immunosuppression stage of sepsis, the release of proinflammatory cytokines is usually reduced, yet the release of anti-inflammatory cytokines is increased or unchanged, which is generally considered as "immunoparalysis" (or endotoxin tolerance). In our study, pro-inflammatory cytokines (IFNG, IL1B, and TNF) were upregulated in the HLA low-risk subgroup, whereas anti-inflammatory cytokines (IL-10) were upregulated in the HLA high-risk subgroup, which is in keeping with the feature of immunocompromise. Additionally, elevated ratios of anti-inflammatory and pro-inflammatory cytokines (e.g., IL-10/TNF) are proposed markers of sepsis-induced immunosuppression and are associated with multiple organ failure (Loisa et al., 2003). Notably, the HLA classifier was significantly positively related to ratios of IL-10/TNF in our study, which implies that the HLA classifier can act as a promising biomarker of sepsis-induced immunoparalysis.

Taken together, according to immune cell infiltration, immune-related pathways, and cytokines level, the HLA classifier could efficiently reflect immunological status, which may help guide immune-modulating agents to achieve immune homeostasis.

In spite of the remarkable results, it is inevitable that limitations also existed in our research. First, though our model, based on multiple transcriptome data, demonstrated impressive performance in early diagnosis, identification of high-risk patients, and recognition of immunosuppression for sepsis, it is not yet suitable for general use prior to validation of external datasets with large sample sizes in prospective cohorts. Second, patients with sepsis included in our analysis were not guaranteed to be free of other diseases. Whereas, the influence of other diseases on our results cannot be fully resolved because the original data set did not offer complete details of other comorbidities/diseases. Third, based on bulk RNA-seq data, the CIBERSORTx deconvolution algorithm and ssGSEA with metagenes may not accurately identify immune cell subpopulations although different methods and different data sets validate each other. It is necessary to use flow cytometry or single-cell RNA-seq methods or fluorescence-activated cell sorting to verify our results. Fourth, no further *in vivo* experiments were conducted to validate these results (hub HLA genes, immune infiltration cells, and pivotal molecular pathways). Loss of function and overexpression studies *in vitro*, as well as in animal models, will help to further identify the exact role of hub HLA genes in the regulation of the inflammatory response and related pathogenic signaling in sepsis.

CONCLUSION

A diagnostic and prognostic model, namely the HLA classifier, was established based on five HLA genes that were closely correlated with responses to hydrocortisone and immunosuppression state, and might facilitate individualized interventions for specific therapy.

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.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The studies involving human tissues samples were reviewed and approved by the Research Ethics Committee of Shunde Hospital, Southern Medical University (The First people's hospital of Shunde), and complied with the Declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZC, RC, JL, and YO conceived and designed the study. ZC, RC, LZ, GL, and QJ drafted the manuscript. JL, ZC, and LW analyzed and interpreted all the data. ZC, LZ, GL,QJ, BY, and YL prepared the figures and tables. ZC, RC, JL, YO, LZ, GL, ZZ, and QJ reviewed and revised the manuscript. All authors have read and approved the manuscript for publication.

FUNDING

Financial support for the research was provided by the Natural Science Foundation of Guangdong Province, China (Grant No. 2018A030313787 to RC) and Scientific Research Start Plan of Shunde Hospital, Southern Medical University (Grant No. SRSP2021010).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2022.870657/full#supplementary-material

Supplementary Figure S1 | Flowchart of dataset selection.

Supplementary Figure S2 | Hub HLA genes selected by Lasso regression analysis and random forest (RF) in GSE63042 datasets. **(A)** The two dotted vertical lines are drawn at the optimal values by minimum criteria (left) and 1—s.e. criteria (right). **(B)** Lasso coefficient profiles of the seven hub HLA genes. A vertical line is drawn at the optimal value by minimum criteria and results in seven non-zero coefficients. **(C)** Distribution diagram of regression tree and error. **(D)** Important variables ranked by IncNodePurity.

Supplementary Figure S3 | Hub HLA genes selected by Lasso regression analysis and random forest (RF) in GSE95233 datasets. **(A)** The two dotted vertical lines are drawn at the optimal values by minimum criteria (left) and 1—s.e. criteria (right). **(B)** Lasso coefficient profiles of the 15 hub HLA genes. A vertical line is drawn at the optimal value by minimum criteria and results in 15 non-zero coefficients. **(C)** Distribution diagram of regression tree and error. **(D)** Important variables ranked by IncNodePurity.

Supplementary Figure S4 | The location of 5 hub HLA genes on 23 chromosomes.

Supplementary Figure S5 | The distribution of mortality rate in different HLA subgroups and the prognostic capacity of the HLA classifier in external validation cohorts. **(A,B)** E-MTAB-4421 datasets. **(C,D)** E-MTAB-4451 datasets. **(E,F)**: E-MTAB-7581 datasets.

Supplementary Figure S6 | The evaluation of the clinical usefulness of the HLA classifier in the discovery cohorts. **(A)** GSE54514 datasets. **(B)** GSE63042 datasets.

Supplementary Figure S7 | The distribution of 22 immune infiltration cells in each sepsis sample from GSE65682 datasets.

Supplementary Figure S8 | The distribution of 22 immune infiltration cells in each sepsis sample from E-MTAB-4421 datasets.

Supplementary Figure S9 | The distribution of 22 immune infiltration cells in each sepsis sample from E-MTAB-4451 datasets.

Supplementary Figure S10 | Bar plot displays differences in pathway activities between different HLA subgroups.

REFERENCES

Baniasadi, A., Rezaeirad, S., Zare, H., and Ghassemi, M. M. (2021). Two-Step Imputation and AdaBoost-Based Classification for Early Prediction of Sepsis on Imbalanced Clinical Data. *Crit. Care Med.* 49 (1), e91–e97. doi:10.1097/ccm.0000000000004705

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Cui, J., Wei, X., Lv, H., Li, Y., Li, P., Chen, Z., et al. (2019). The Clinical Efficacy of Intravenous IgM-Enriched Immunoglobulin (Pentaglobin) in Sepsis or Septic Shock: A Meta-Analysis with Trial Sequential Analysis. *Ann. Intensive Care* 9 (1), 27. doi:10.1186/s13613-019-0501-3

Davenport, E. E., Burnham, K. L., Radhakrishnan, J., Humburg, P., Hutton, P., Mills, T. C., et al. (2016). Genomic Landscape of the Individual Host Response and Outcomes in Sepsis: A Prospective Cohort Study. *Lancet Respir. Med.* 4 (4), 259–271. doi:10.1016/s2213-2600(16)00046-1

Dong, Y., Liu, Y., Bai, H., and Jiao, S. (2019). Systematic Assessment of the Clinicopathological Prognostic Significance of Tissue Cytokine Expression for Lung Adenocarcinoma Based on Integrative Analysis of TCGA Data. *Sci. Rep.* 9 (1), 6301. doi:10.1038/s41598-019-42345-0

Ferrer, R., Martin-Loeches, I., Phillips, G., Osborn, T. M., Townsend, S., Dellinger, R. P., et al. (2014). Empiric Antibiotic Treatment Reduces Mortality in Severe Sepsis and Septic Shock from the First Hour: Results from a Guideline-Based Performance Improvement Program. *Crit. Care Med.* 42 (8), 1749–1755. doi:10.1097/ccm.0000000000000330

Fleischmann, C., Scherag, A., Adhikari, N. K. J., Hartog, C. S., Tsaganos, T., Schlattmann, P., et al. (2016). Assessment of Global Incidence and Mortality of Hospital-Treated Sepsis. Current Estimates and Limitations. *Am. J. Respir. Crit. Care Med.* 193 (3), 259–272. doi:10.1164/rccm.201504-0781oc

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinform.* 14, 7. doi:10.1186/1471-2105-14-7

Henriquez-Camacho, C., and Losa, J. (2014). Biomarkers for Sepsis. *Biomed. Res. Int.* 2014, 547818. doi:10.1155/2014/547818

Hotchkiss, R. S., Monneret, G., and Payen, D. (2013). Sepsis-induced Immunosuppression: From Cellular Dysfunctions to Immunotherapy. *Nat. Rev. Immunol.* 13 (12), 862–874. doi:10.1038/nri3552

Hotchkiss, R. S., Moldawer, L. L., Opal, S. M., Reinhart, K., Turnbull, I. R., and Vincent, J.-L. (2016). Sepsis and Septic Shock. *Nat. Rev. Dis. Prim.* 2, 16045. doi:10.1038/nrdp.2016.45

Jensen, J.-U., and Bouadma, L. (2016). Why Biomarkers Failed in Sepsis. *Intensive Care Med.* 42 (12), 2049–2051. doi:10.1007/s00134-016-4531-0

Kjaergaard, A. G., Nielsen, J. S., Tønnesen, E., and Krog, J. (2015). Expression of NK Cell and Monocyte Receptors in Critically Ill Patients - Potential Biomarkers of Sepsis. *Scand. J. Immunol.* 81 (4), 249–258. doi:10.1111/sji.12272

Koşaloğlu-Yalçın, Z., Lanka, M., Frentzen, A., Logandha Ramamoorthy Premlal, A., Sidney, J., Vaughan, K., et al. (2018). Predicting T Cell Recognition of MHC Class I Restricted Neoepitopes. *Oncoimmunology* 7 (11), e1492508. doi:10.1080/2162402X.2018.1492508

Kumar, V. (2018). T Cells and Their Immunometabolism: A Novel Way to Understanding Sepsis Immunopathogenesis and Future Therapeutics. *Eur. J. Cell. Biol.* 97 (6), 379–392. doi:10.1016/j.ejcb.2018.05.001

Loisa, P., Rinne, T., Laine, S., Hurme, M., and Kaukinen, S. (2003). Anti-inflammatory Cytokine Response and the Development of Multiple Organ Failure in Severe Sepsis. *Acta Anaesthesiol. Scand.* 47 (3), 319–325. doi:10.1034/j.1399-6576.2003.00004.x

Monneret, G., Lepape, A., Voirin, N., Bohé, J., Venet, F., Debard, A.-L., et al. (2006). Persisting Low Monocyte Human Leukocyte Antigen-DR Expression Predicts Mortality in Septic Shock. *Intensive Care Med.* 32 (8), 1175–1183. doi:10.1007/s00134-006-0204-8

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., et al. (2020). Global, Regional, and National Sepsis Incidence and Mortality, 1990-2017: Analysis for the Global Burden of Disease Study. *Lancet* 395 (10219), 200–211. doi:10.1016/s0140-6736(19)32989-7

Schaafsma, E., Fugle, C. M., Wang, X., and Cheng, C. (2021). Pan-cancer Association of HLA Gene Expression with Cancer Prognosis and Immunotherapy Efficacy. *Br. J. Cancer* 125 (3), 422–432. doi:10.1038/s41416-021-01400-2

Schrijver, I. T., Théroude, C., and Roger, T. (2019). Myeloid-Derived Suppressor Cells in Sepsis. *Front. Immunol.* 10, 327. doi:10.3389/fimmu.2019.00327

Simpson, S. Q. (2017). Diagnosing Sepsis: A Step Forward, and Possibly a Step Back. *Ann. Transl. Med.* 5 (3), 55. doi:10.21037/atm.2017.01.06

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315 (8), 801–810. doi:10.1001/jama.2016.0287

Steinhagen, F., Schmidt, S. V., Schewe, J.-C., Peukert, K., Klinman, D. M., and Bode, C. (2020). Immunotherapy in Sepsis - Brake or Accelerate? *Pharmacol. Ther.* 208, 107476. doi:10.1016/j.pharmthera.2020.107476

Vickers, A. J., and Elkin, E. B. (2006). Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* 26 (6), 565–574. doi:10.1177/0272989x06295361

Wu, H.-P., Chung, K., Lin, C.-Y., Jiang, B.-Y., Chuang, D.-Y., and Liu, Y.-C. (2013). Associations of T Helper 1, 2, 17 and Regulatory T Lymphocytes with Mortality in Severe Sepsis. *Inflamm. Res.* 62 (8), 751–763. doi:10.1007/s00011-013-0630-3

Zhai, J., Qi, A., Zhang, Y., Jiao, L., Liu, Y., and Shou, S. (2020). Bioinformatics Analysis for Multiple Gene Expression Profiles in Sepsis. *Med. Sci. Monit.* 26, e920818. doi:10.12659/MSM.920818

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, Chen, Ou, Lu, Jiang, Liu, Wang, Liu, Zhou, Yang and Zuo. 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 Molecular Biosciences

ORIGINAL RESEARCH
published: 13 June 2022
doi: 10.3389/fmolb.2022.923584

Identification and Validation of lncRNA-AC087588.2 in Lung Adenocarcinoma: A Novel Prognostic and Diagnostic Indicator

Xiulin Jiang[1†], Xi Chen[2†], Jishu Guo[3†], Fan Zhou[4†], Jun Pu[2], Luciano Mutti[5] and Xiaoqun Niu[6*]*

[1]Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing, China, [2]Department of Neurosurgery, The Second Affiliated Hospital of Kunming Medical University, Kunming, China, [3]Institute for Ecological Research and Pollution Control of Plateau Lakes, School of Ecology and Environmental Science, Yunnan University, Kunming, China, [4]Hematology and Rheumatology Department, The Pu'er People's Hospital, Puer, China, [5]Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, PA, United States, [6]Department of Respiratory Medicine, Second Hospital of Kunming Medical University, Kunming, China

OPEN ACCESS

Edited by:
William C. Cho, QEH, Hong Kong SAR, China

Reviewed by:
Jianyun Dong, First Affiliated Hospital of Southern University of Science and Technology, China
Juncheng Zhang, The First Affiliated Hospital of Sun Yat-sen University, China

***Correspondence:**
Luciano Mutti
luciano.mutti@temple.edu
Xiaoqun Niu
673409256@qq.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *19 April 2022*
Accepted: *04 May 2022*
Published: *13 June 2022*

Citation:
Jiang X, Chen X, Guo J, Zhou F, Pu J, Mutti L and Niu X (2022) Identification and Validation of lncRNA-AC087588.2 in Lung Adenocarcinoma: A Novel Prognostic and Diagnostic Indicator. Front. Mol. Biosci. 9:923584. doi: 10.3389/fmolb.2022.923584

Lung adenocarcinoma (LUAD) is the most common histological lung cancer, and it is the leading cause of cancer-related deaths worldwide. Long non-coding RNAs (lncRNAs) have been implicated in the initiation and progression of various cancers. LncRNA-AC087588.2 (ENSG00000274976) is a novel lncRNA that is abnormally expressed in diverse cancer types, including LUAD. However, the clinical significance, prognostic value, diagnostic value, immune role, and the potential biological function of AC087588.2 LUAD remain elusive. In this study, we found that AC087588.2 was upregulated and associated with a poor prognosis in LUAD. In addition, univariate and multivariate Cox regression analysis indicated that AC087588.2 could be an independent prognostic factor for LUAD. Functionally, the knockdown of AC087588.2 restrained LUAD cell proliferation and migration *in vitro*. Finally, we constructed a ceRNA network that included hsa-miR-30a-5p and four mRNAs (ANLN, POLR3G, EHBP1, and ERO1A) specific to AC087588.2 in LUAD. The Kaplan–Meier survival analysis showed that lower expression of hsa-miR-30a-5p and higher expression of ANLN, POLR3G, EHBP1, and ERO1A were associated with adverse clinical outcomes in patients with LUAD. This finding provided a comprehensive view of the AC087588.2-mediated ceRNA network in LUAD, thereby highlighting its potential role in the diagnosis and prognosis of LUAD.

Keywords: lncRNA, lung adenocarcinoma, prognosis biomarker, immune infiltration, ceRNA, cell proliferation, cell migration

INTRODUCTION

Lung cancer includes small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC). NSCLC includes lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and large-cell lung carcinoma. The NSCLC cancer accounts for approximately 85% of all cases (Molina et al., 2008). Although the treatment of LUAD has improved, the new LUAD pathogenesis and

noninvasive diagnostic biomarkers are still needed. Therefore, the discovery of potential key prognostic markers with more characteristics and value will help early prediction and treatment of LUAD at the molecular level.

As a newly discovered non-coding RNA, the length of lncRNA usually exceeds 200 nt without protein-coding capacity (Deng et al., 2020). Emerging evidence demonstrates that lncRNA is involved in various physiological and pathological processes, including lung cancer (Lu et al., 2017). For instance, Wu et al. found that linc00673 was highly associated with poor prognosis in NSCLC. A further study showed that, by sponging miR-150-5p and upregulating ZEB1 expression, linc00673 promotes NSCLC proliferation, migration, invasion, and epithelial-mesenchymal transition (Lu et al., 2017). LncRNA ZEB1-AS1 was found to be correlated with the EMT process and adverse prognosis in LUAD (Li et al., 2016). Shen et al. found that lncRNA FEZF1-AS1 was increased in NSCLC tissues compared with adjacent normal tissues. Mechanism research indicated that, by regulating the wnt pathway, FEZF1-AS1 inhibited the EMT of NSCLC (He et al., 2017). We recently used a new method, cross-value association analysis (CVAA), to analyze the TCGA-LUAD dataset, and identified numerous new differentially expressed genes (DEGs), including lncRNA-AC087588. However, the clinical value and biological function of AC087588.2 in LUAD have not been explored.

In the present study, we use various databases to analyze AC087588.2 expression, clinical significance, prognostic value, diagnostic value, and immune infiltration and determine the potential oncogenic function in LUAD. Meanwhile, qRT-PCR, CCK8, wound healing, and transwell assays were employed to determine the potential biological function of AC087588.2 in LUAD progression.

MATERIALS AND METHODS

TCGA Datasets

Transcription and clinical information of LUAD was downloaded from TCGA (https://portal.gdc.com) (Tomczak et al., 2015). RNA-seq gene expression data of workflow type FPKM were transformed into TPM format and log2 transformation for further study. The timeROC analysis was used to compare the predictive accuracy of the AC087588.2 gene in LAUD. Clinical information of the LUAD patients consisted of the pathological stage, TNM stage, smoker, OS, DSS, and PFS.

Cox Regression Analysis and Kaplan–Meier Survival Analysis

We utilized Cox regression analysis to examine the correlation between AC087588.2 expression and overall survival and disease-specific survival of patients using the TCGA databases. The Kaplan–Meier method was used to assess the difference between high- and low-risk groups based on the best separation of AC087588.2 expression employing R packages of survminer and survival.

Gene Set Enrichment Analysis

In the present research, we utilized the LinkeDomics database (http://www.linkedomics.org/login.php) to obtain the co-expression genes of AC087588.2 in LUAD. KEGG and GO were employed to assess the potential functions of AC087588.2 in LUAD with the R package ClusterProfiler (Yu et al., 2012). GSEA software was used to explore the potential signaling pathway involved by AC087588.2 in LUAD (Subramanian et al., 2005; Yu et al., 2012; Vasaikar et al., 2018).

The Target Gene of miR-30a-5p Predicted by Various Databases

StarBase (https://starbase.sysu.edu.cn/) is a database that includes the miRNA-ceRNA, miRNA-ncRNA, and protein–RNA interaction networks from large-scale CLIP-Seq data. In this manuscript, we used StarBase to predict the target gene of miR-30a-5p. StarBase was also utilized to determine the relationship between miR-30a-5p expression and AC087588.2 in LUAD (Cho et al., 2013; Dweep et al., 2014; Li et al., 2014; Chen and Wang, 2020).

Immune Infiltration Analysis by ssGSEA

In this study, the ssGSEA method was employed to analyze the association between AC087588.2 expression and the infiltration of 24 tumor-infiltrating lymphocytes (TILs) in LUAD (Bindea et al., 2013)

Cell Culture Conditions

Lung cancer cell lines, including one human normal bronchial epithelial cell (BEAS-2B) and three human lung adenocarcinoma cells (H1975, SPC-A1, H1299, and A549 cells), were purchased from the Chinese Academy of Sciences Cell Bank (CASCB, China) and cultured in RPMI 1640 medium (Corning) including 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin at 37°C an atmosphere containing 95% air and 5% CO_2.

siRNA and Transfection

SiRNA for AC087588.2 and the matched negative controls were designed and synthesized by RiboBio (Guangzhou, China). Lipofectamine™ 3000 Reagent (Invitrogen, United States) was used to transfect siRNA according to the manufacturer's instructions. The primer used in this study is as follows: AC087588.2 siRNA#1: GCCTTGGTCATGAAACGATTA.

Quantitative Real-Time PCR

The qRT-PCR assay was performed as documented (Pan et al., 2020). The primer sequences are as follows: AC087588.2 -F: GCA CTTACTTTATAGCAGCAA, AC087588.2 -R: ATAAATATG GTTTCTCAAGT; β-actin-F: CTTCGCGGGCGACGAT, β-actin-R: CCATAGGAATCCTTCTGACC. The expression quantification was obtained with the 2−ΔΔCt method.

Cell Migration Assay

For the transwell migration assay, 2.5×10^4 cells/well in 100 μL serum-free medium were plated in a 24-well plate chamber insert,

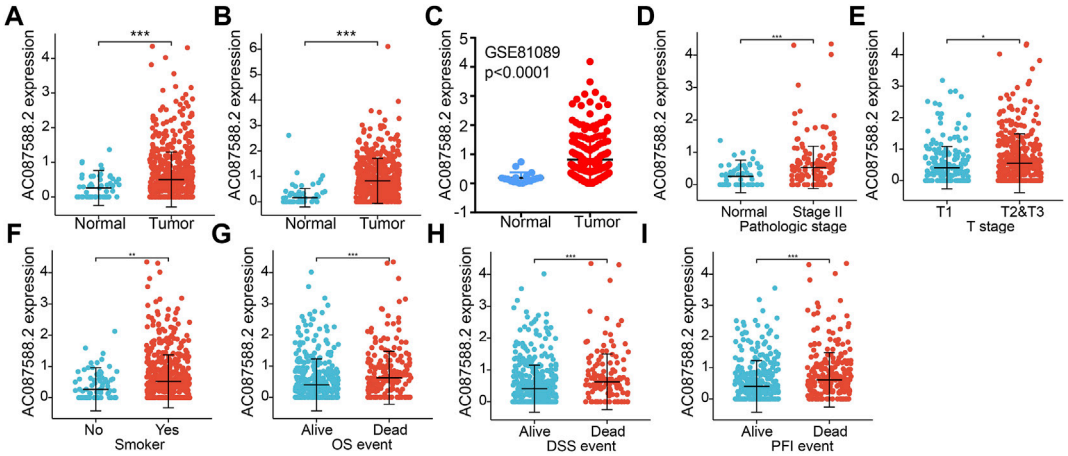


FIGURE 1 | AC087588.2 was overexpressed in LUAD. **(A–C)** AC087588.2 was overexpressed in LUAD and LUSC examined by TCGA and GEO datasets. **(D–I)** Correlation between AC087588.2 expression and clinical parameters, including the pathological stage, T stage, smoking, overall survival event, disease-specific survival event, and progression-free survival event. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

and the lower chamber was filled with 10% FBS. After incubation for 24 h, cells were fixed with 4% paraformaldehyde, washed, and then stained with 0.5% crystal violet for further pictures captured.

CCK8 Assay

We seeded cells in 96-well plates at 2.5×10^3 per well in 100 µl of complete medium and 10 µl of CCK-8 reagent (RiboBio, Guangzhou, China) for 1 h each day after 3 days of culture. We then used a microplate to measure the absorbance of each well at 450 nm. Each sample was evaluated in triplicate.

Statistical Analyses

The significance of the data between two experimental groups was determined by Student's *t*-test, and multiple group comparisons were analyzed by one-way ANOVA. $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) were significant.

RESULTS

Differential Expression of AC087588.2 in LUAD Patients

To examine AC087588.2 expression in LUAD, we analyzed AC087588.2 expression data in TCGA-LUAD and uncovered that AC087588.2 was upregulated in 535 tumor tissues compared to 59 normal lung tissues in LUAD and upregulated in 502 tumor tissues compared to 49 normal lung tissues in LUSC (**Figures 1A,B**). Further analyses of one dataset of GEO obtained similar results (**Figure 1C**). To determine the potential function of AC087588.2 in the development of LUAD, we analyzed the correlation between AC087588.2 expression and various clinical features. The analysis indicated that AC087588.2 was significantly associated with advanced pathological stage, T stage, smoking, OS event, DSS event, and PFS event (**Figures 1D–I**).

AC087588.2 Modulates Cell Proliferation and Migration of LUAD Cells

To explore the biological role of AC087588.2 in LUAD, we found that AC087588.2 expression was significantly upregulated in H1975, SPC-A1, H1299, and A549 cell lines (**Figure 2A**). Moreover, specific siRNA for AC087588.2 was used to knock down the AC087588.2 expression (**Figures 2B,C**). It was shown that the depletion of AC087588.2 inhibited the proliferative capacity of LUAD cells (**Figures 2D,E**). Moreover, transwell and wound healing assay confirmed that AC087588.2 knockdown inhibited the cell migration of LUAD (**Figures 2F–I**). These data indicated that AC087588.2 is functionally important in regulating LUAD progression.

Association Between AC087588.2 Expression and Clinical Outcome in LUAD

Next, to determine the correlation between AC087588.2 expression and clinical outcome of LUAD patients, the TCGA-LUAD datasets were employed. Based on the median expression of AC087588.2 in LUAD, the expression level of AC087588.2 in LUAD patients was divided into two groups with high and low expression. Results suggested that upregulation of AC087588.2 was significantly associated with poor OS, DSS, and PFS in patients with LUAD (**Figures 3A–C**). Based on the time-dependent ROC, the AC087588.2 expression level had a relatively good performance in predicting the 1-, 3-, and 5-year OS, DSS, and PFS in LUAD patients (**Figures 3D–F**). We also utilized GEO datasets validation the prognosis of AC087588.2 in lung cancer and obtained the same results (**Figures 3G–I**). Moreover, an overall survival analysis was conducted to determine the prognostic value of AC087588.2 in different subgroups of LUAD patients stratified by stages II and III, stage I, stages I and III, T1 and T2, T3 and T4, N0 and N1, N2 and N3, M0, primary: CR, R0, female, male, age >65, and smoker. Results suggested that increased AC087588.2 level is associated with poor prognosis in patients with LUAD (**Figures 4A–D**).

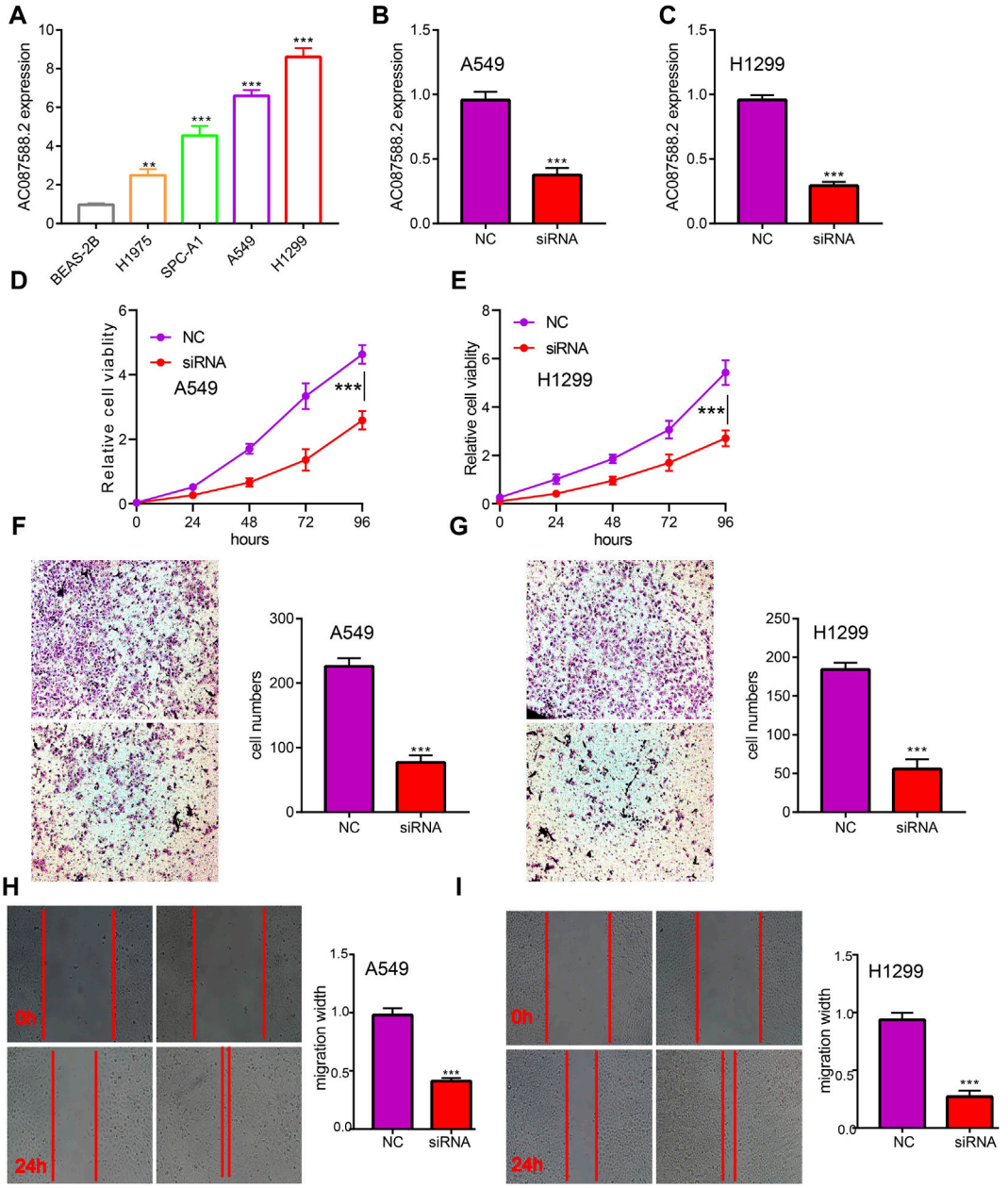


FIGURE 2 | AC087588.2 modulates LUAD cell proliferation and migration *in vitro*. **(A)** The expression of AC087588.2 in LUAD cell lines examined by qPCR. **(B–C)** qPCR assay examined the knockdown efficiency of AC087588.2 in A549 and H1299 cells **(D–I)** AC087588.2 knockdown inhibited cell proliferation and migration examined by CCK8, transwell and wound healing assay. $^{**}p < 0.01$, and $^{***}p < 0.001$. NC = negative control, siRNA = AC087588.2 siRNA.

AC087588.2 as an Independent Risk Factor

To determine whether AC087588.2 is an independent factor for the prognosis of LUAD, we conducted the univariate and multivariate Cox regression analyses based on the TCGA-LUAD. Results suggested that the AC087588.2 expression, pathological stage, and N stage could be independent risk factors for LUAD (**Figures 5A,B**). Then, we constructed the nomograms using the above independent prognosis factors (N stage and AC087588.2 expression) to predict 1-, 3-, and 5-year OS, DSS, and PFS of each LUAD patient (**Figures 5C–E**). The calibration plot of survival suggested that this nomogram could predicate OS, DSS, and PFS (**Figures 5F–H**).

KEGG Enrichment Analysis for AC087588.2 in LUAD

To examine the potential biological functions of AC087588.2 in LUAD, we performed KEGG enrichment analysis on genes that were significantly positively correlated with AC087588.2 expression based on the TCGA-LAUD dataset (**Figures 6A,B**). The GO

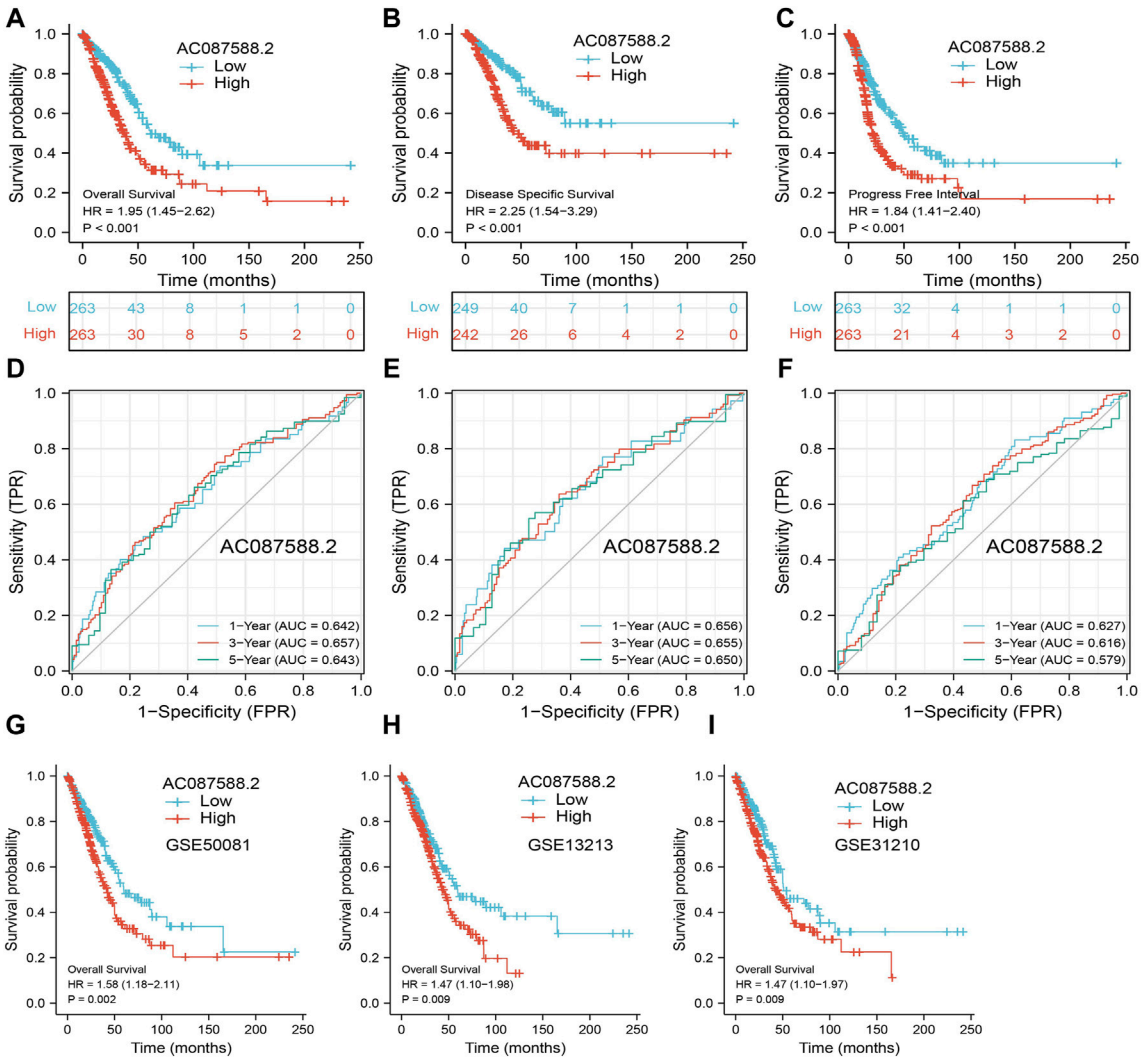


FIGURE 3 | ROC and Kaplan–Meier curves of AC087588.2. **(A–C)** OS, DSS, and PFS of AC087588.2 in LUAD determined using the TCGA-LUAD dataset. **(D–F)** ROC curve of AC087588.2 in LUAD determined using the TCGA-LUAD dataset. **(G–I)** Validation of overall survival of AC087588.2 in lung cancer by GEO datasets.

functional analyses suggested that AC087588.2 was mainly involved in the regulation of DNA metabolic process, chromosome segregation, and proteasomal protein catabolic process regulation of chromosome organization (**Figure 6C**). Meanwhile, the KEGG pathway analyses confirmed that AC087588.2-related pathways involve the PI3K–Akt signaling pathway, MAPK signaling pathway, JAK–STAT signaling pathway, and TNF signaling pathway (**Figure 6D**). Gene set enrichment analysis (GSEA) also showed that AC087588.2 was mainly involved in the cell apoptosis, EMT, G2/M checkpoint, glycolysis, p53, MYC targets, and oxidative phosphorylation (**Figures 7A–D**).

Correlation Between AC087588.2 Expression and Immune Infiltration

Immune infiltration has a crucial role in LUAD development [25]. Therefore, we explore the correlation between AC087588.2 expression and the infiltration levels of 24 types of immune cells in LUAD using the ssGSEA method. Results suggested that the AC087588.2 expression in LUAD was positively related to the infiltration of Th2 cells and negatively associated with the abundance of B cells, Th17 cells, macrophages, DC, eosinophils, iDC, TFH, and mast cells in LUAD (**Figures 8A–C**).

Given that immune checkpoints play a crucial role in tumor immunosuppression, we analyzed the correlation between AC087588.2 expression and that of the immune checkpoint-related genes of *CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *PDCD1LG2*, *TIGIT*, and *SIGLEC15* in LUAD using Pearson's correlation analysis. AC087588.2 expression was significantly positively correlated with the expression of *CD274*, *CTLA4*, *LAG3*, *TIGIT*, and *SIGLEC15* in this analysis (**Figure 8D**). These results confirmed that AC087588.2 played a crucial role in immune infiltration in LUAD.

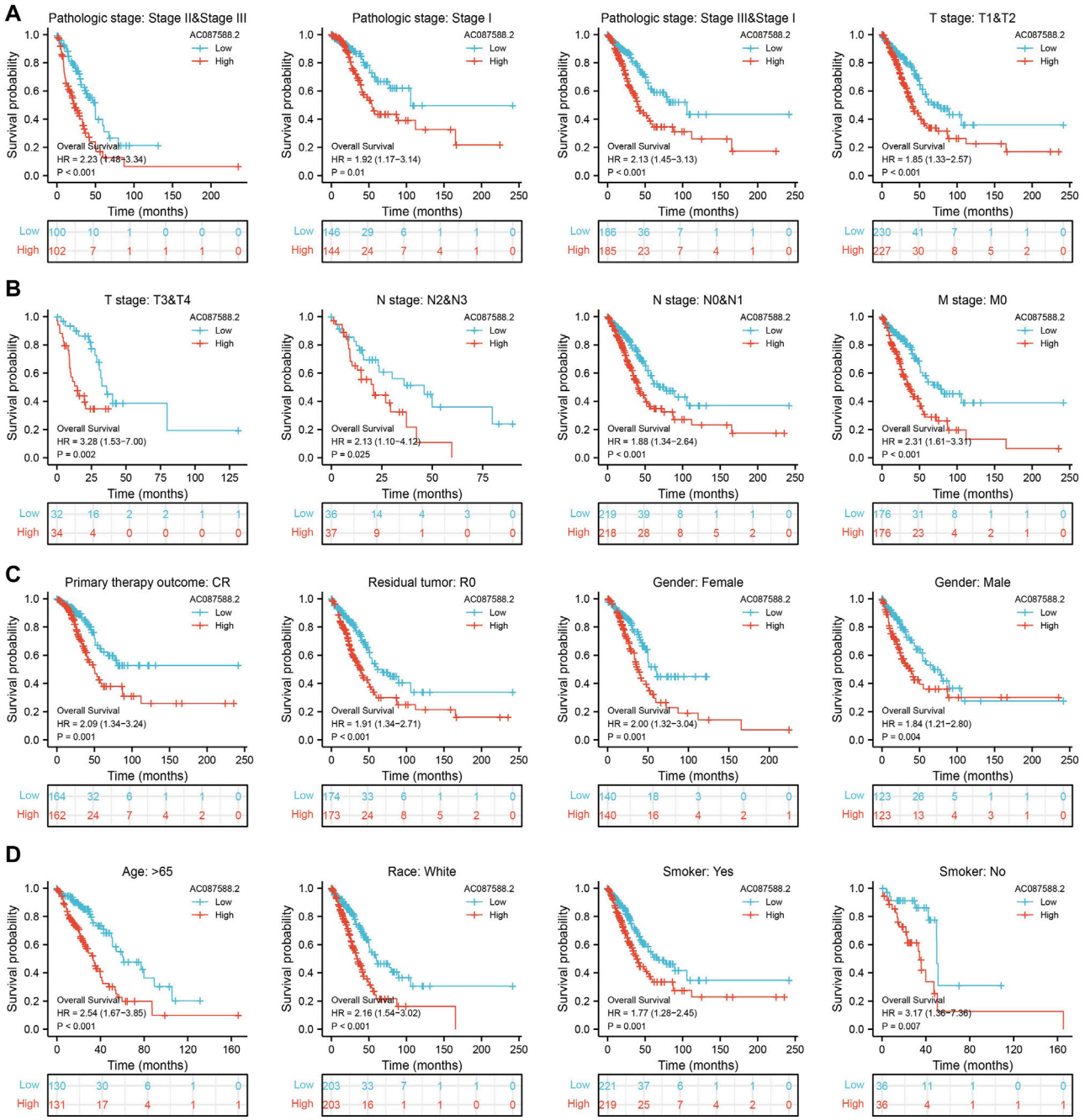
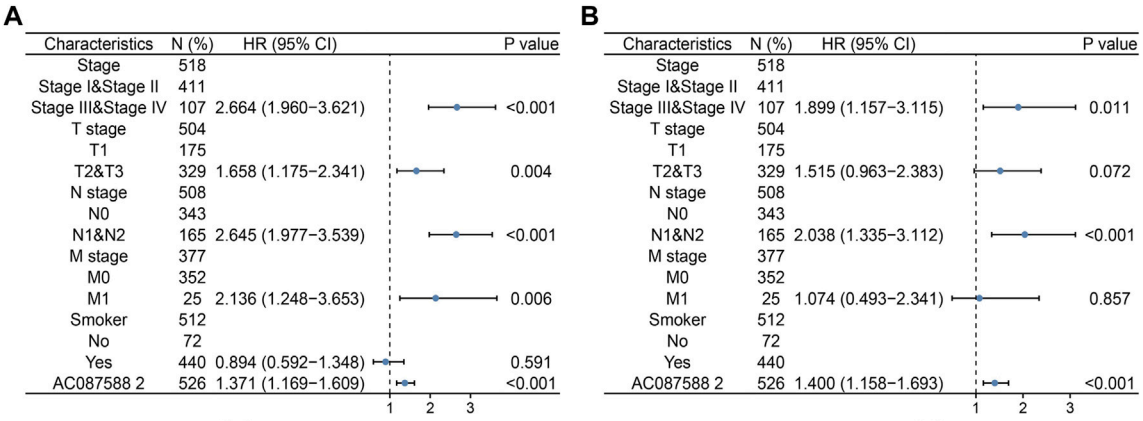
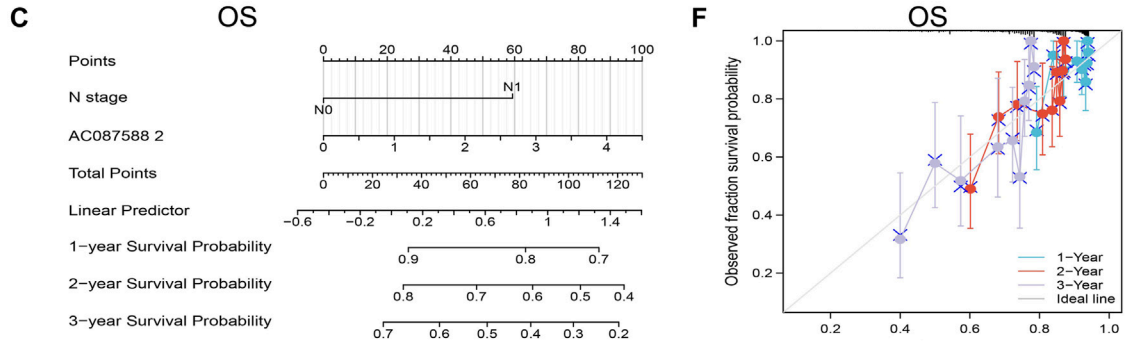


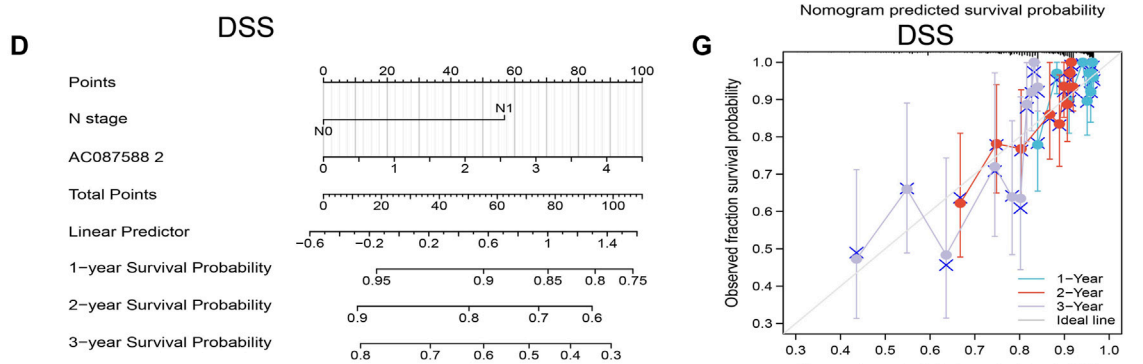
FIGURE 4 | Overall survival of AC087588.2 based on the diverse subgroup. **(A–D)** Overall survival of AC087588.2 based on the diverse subgroup, including stages II and III, stage I, stages I and III, T1 and T2, T3 and T4, N0 and N1, N2 and N3, M0, primary: CR, R0, female, male, age >65, smoker.

AC087588.2-Related miRNA–mRNA Network in LUAD

To further explore the AC087588.2-mediated downstream regulatory mechanism involved in LUAD progression, we used the Annolnc2 (http://annolnc.gao-lab.org/) database to predict the potential miRNAs binding with AC087588.2. We obtained 10 miRNAs (**Figure 9A**) (Ke et al., 2020). Based on the competitive endogenous RNAs theory, lncRNA should be positively correlated with mRNA and negatively correlated with miRNA. Among all the 10 miRNAs, only miRNA-30a-5p negatively correlated with AC087588.2 in LUAD (**Figure 9B**). Moreover, we found that has-miR-30a-5p had low expression in LUAD, which correlated with poor prognosis in patients with LUAD (**Figures 9C,D**). ROC curve analysis showed that the AUC value of has-miR-30a-5p is 0. 824 (**Figure 9E**). Therefore, we selected has-miR-30a-5p to conduct downstream analysis.







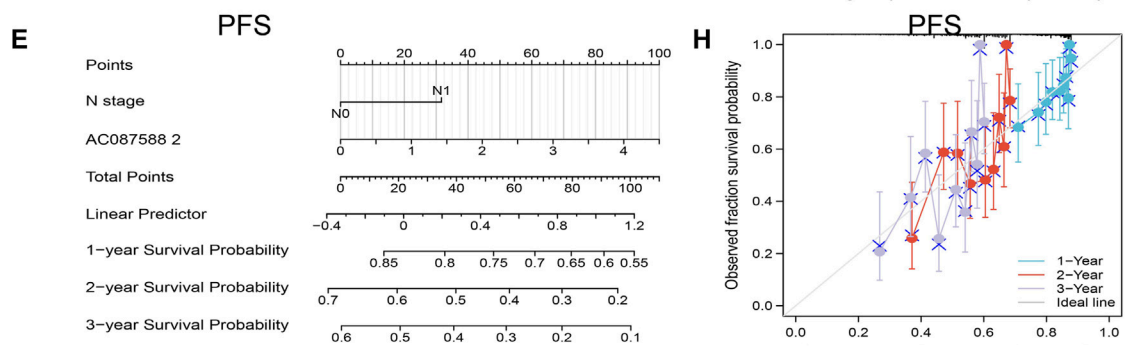


FIGURE 5 | Univariate and multivariate Cox regression analysis. **(A,B)** Univariate and multivariate Cox regression analysis in LUAD. **(C–E)** The nomogram was developed by integrating the AC087588.2 expression and T stage in the TCGA databases. **(F–H)** Predicting abilities of the nomogram.

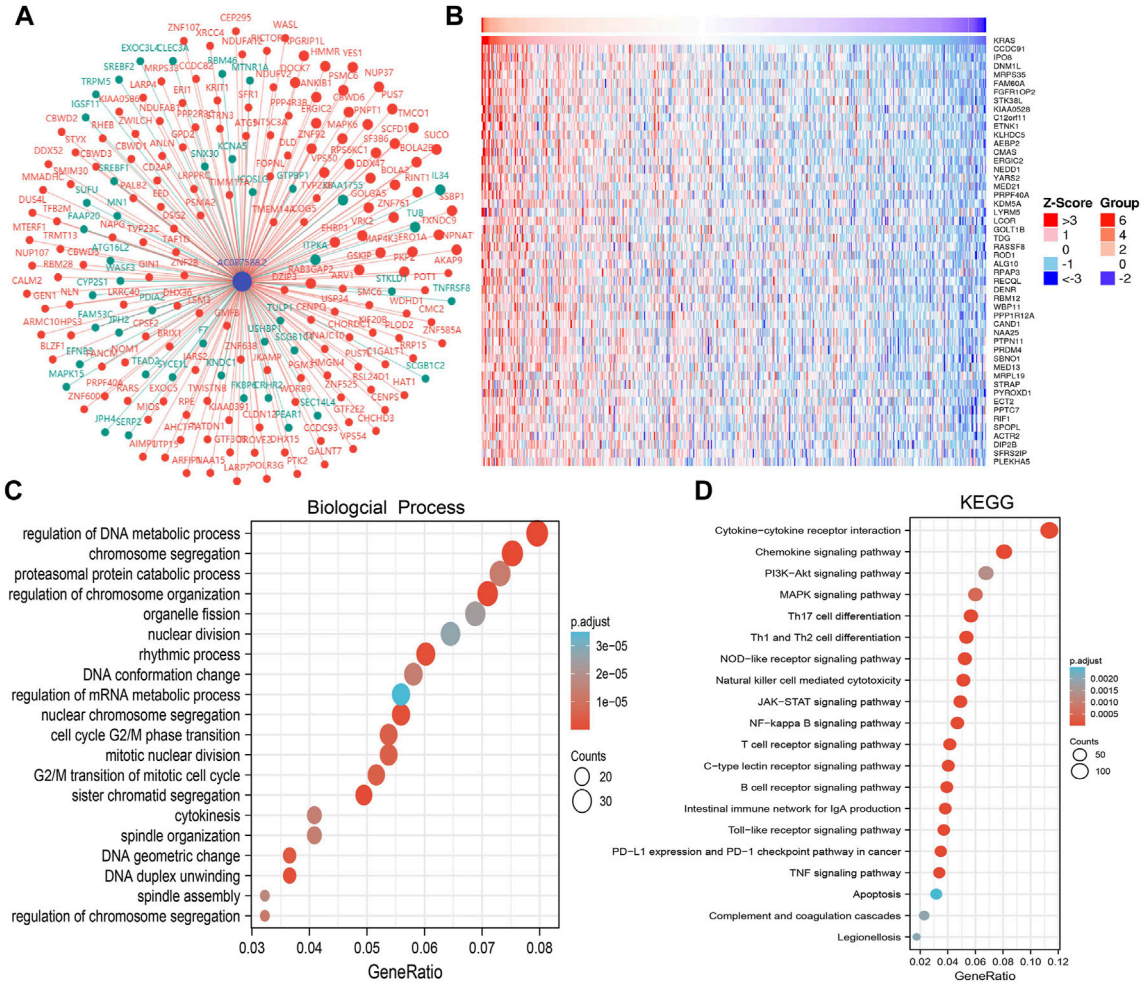


FIGURE 6 | KEGG enrichment analysis of AC087588.2. **(A,B)** Gene–gene interaction network of AC087588.2 in LUAD. **(C,D)** GO and KEGG enrichment analysis of AC087588.2 in LUAD.

Identification of the Potential Downstream Target of AC087588.2/miR-30a-5p in LUAD

We further investigated the target genes of miR-30a-5p that play critical roles in the progression of LUAD. First, we predicted the target in StarBase, miRDB, miRWalk, and miRGator (Cho et al., 2013; Dweep et al., 2014; Li et al., 2014; Chen and Wang, 2020). According to the prediction of target genes, we found that only four genes (ANLN, POLR3G, EHBP1, and ERO1A) were negatively correlated with the miR-30a-5p expression in LUAD (**Figure 9F**). Importantly, the expression levels of ANLN, POLR3G, EHBP1, and ERO1A were positively correlated with those of AC087588.2in LUAD (**Figure 9G**). Furthermore, we employed the TCGA to explore the expression level and prognosis in LUAD. We found that the expression levels of ANLN, POLR3G, EHBP1, and ERO1A were significantly increased in LUAD and associated with OS, DSS, and PFS in patients with LUAD (**Figure 9H**; **10A–C**). ROC curve was utilized to examine the diagnostic value of ANLN, POLR3G, EHBP1, and ERO1A in LUAD, the AUC (area under the curve) of which were 0.978, 0.669, 0.629, and 0.927, respectively (**Figure 10D**), suggesting that ANLN, POLR3G, EHBP1, and ERO1A were a potential prognostic and diagnostic biomarker in LUAD. Finally, we used the ssGSEA method to determine the correlations between ANLN, POLR3G, EHBP1, and ERO1A and 24 types of tumor-infiltrating immune cells. Results confirmed that the expression levels of ANLN, POLR3G, EHBP1, and ERO1A were positively correlated with the cell infiltrating of Th2 cells in LUAD (**Figure 10E**).

DISCUSSION

At present, common treatments for lung cancer mainly include surgical resection, radiation-chemotherapy, and immunotherapy, but the therapeutic effect is not ideal. Consequently, it is urgent to identify novel cancer biomarkers and understand the potential molecular

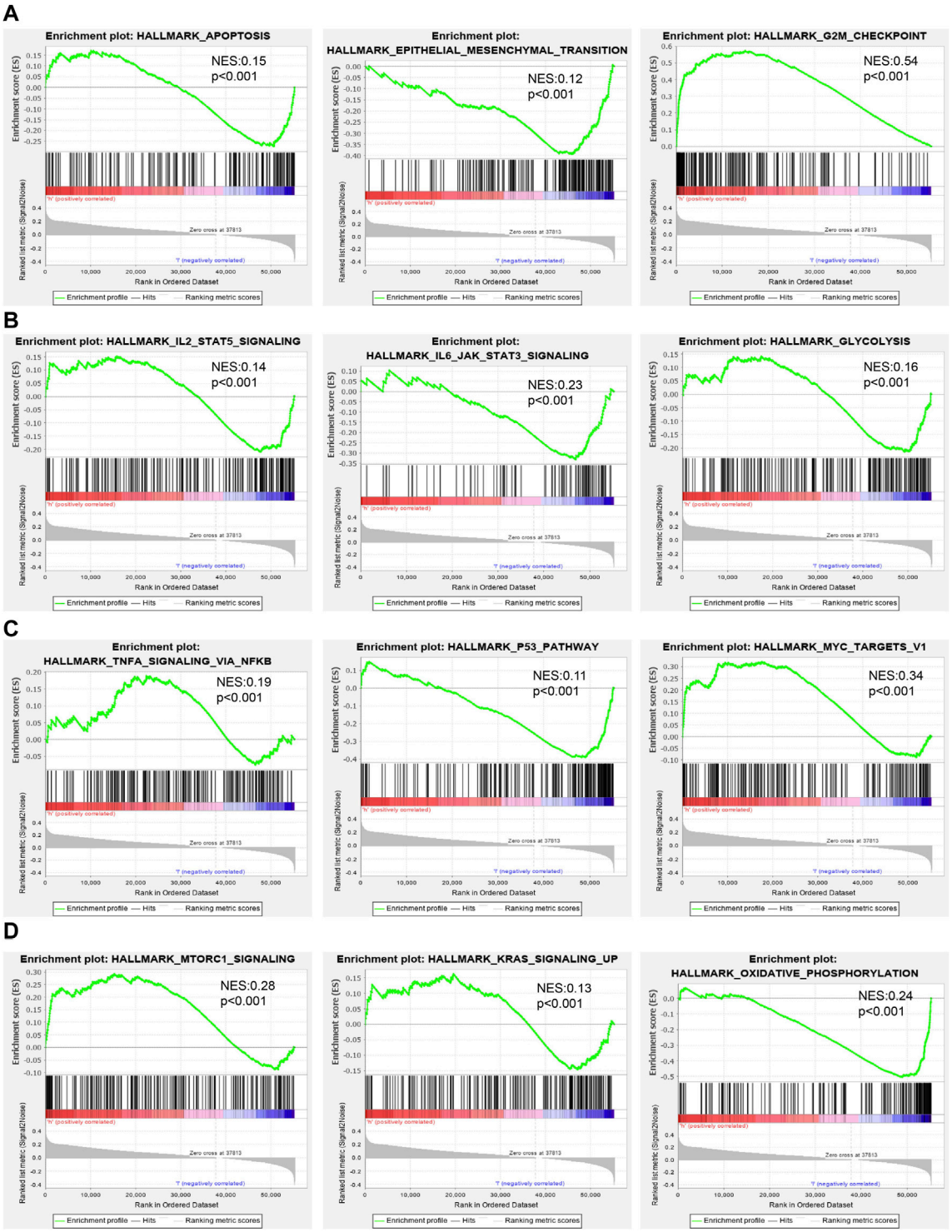


FIGURE 7 | GSEA identification of AC087588.2-related signaling pathways. **(A–D)** Identification of AC087588.2-related signaling pathways by GSEA software.

mechanisms involved in LUAD initiation and progression. It has been confirmed that lncRNAs play an important role in modulating cell proliferation, cell apoptosis, and cancer progression (Dong et al., 2018).

Increasing evidence demonstrated the functional and clinical role of lncRNAs involved in the progression (Jia et al., 2019; Xue et al., 2021). For example, it has been shown that the increased AC079630.4 expression is related to the progression and

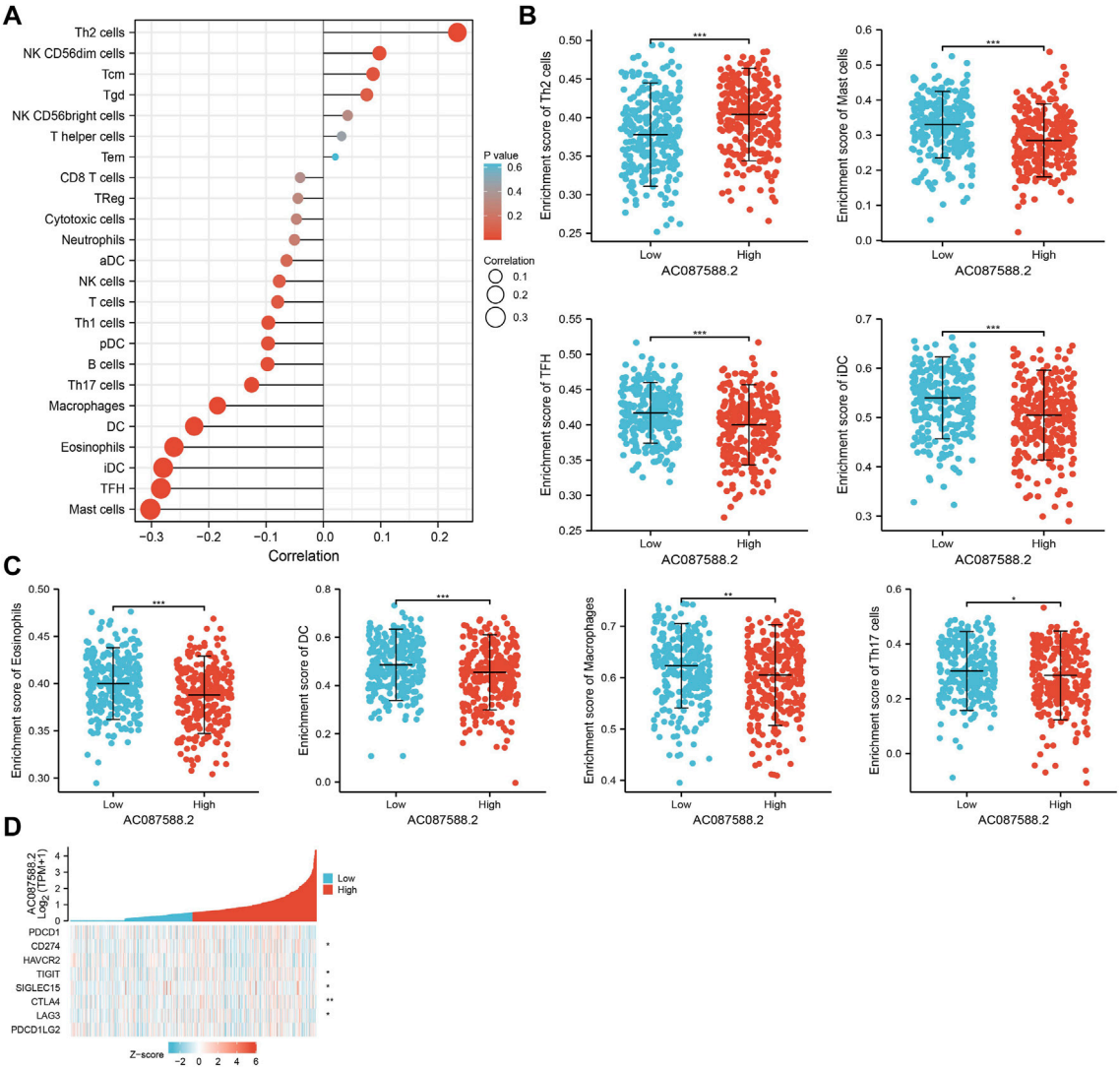


FIGURE 8 | Association between AC087588.2 expression and immune cell infiltration in LUAD. **(A–C)** Correlation between AC087588.2 expression and immune cell infiltration in LUAD. **(D)** Correlation between AC087588.2 expression and immune checkpoint-related genes in LUAD. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

prognosis in lung cancer (Wang et al., 2021). Previous studies indicated that LncRNAs have clinical predictor value in several tumors (Chao and Zhou, 2019). For instance, Wang et al. found that linc8087 was downregulated in NSCLC and its lower expression was related to adverse prognosis in patients with NSCLC (Qi et al., 2021).

In the current study, we uncovered that AC087588.2 was overexpressed in LUAD, and its higher expression was correlated with adverse OS, DSS, and PFS in patients with LUAD. Additionally, ROC curve analysis confirmed that the AUC value of AC087588.2 is 0.888. Results suggested that AC087588.2 could serve as a sensitive indicator to predict the prognosis of the patients, indicating the value of AC087588.2 as a prognostic biomarker for LUAD. Multivariate analysis indicated that the AC087588.2 expression was an independent prognostic indicator for the OS, DSS, and PFS of LUAD patients.

Previous studies reported that lncRNA plays an important role in the EMT and cell cycle (Jia et al., 2019; Geng et al., 2021). For example, it has been confirmed that lncRNA-JPX modulates cell proliferation and migration by sponging miR-33a-5p to increase Twist1 expression (Pan et al., 2020). In this study, we found that AC087588.2 was mainly involved in the cell apoptosis, EMT, G2/M checkpoint, glycolysis, p53, MYC targets, and oxidative phosphorylation.

It has been confirmed that lncRNA plays a central role in facilitating tumor progression and immune escape (Zhang et al., 2020). For example, lncRNA GATA3-AS1 promoted BRCA immune escape by stabilizing PD-L1 (Zhang et al., 2020). In this finding, we demonstrated that the AC087588.2 expression in LUAD was positively associated with the infiltration of Th2 cells and negatively correlated with the abundance of B cells, Th17 cells, macrophages, DC, eosinophils, iDC, TFH, and mast cells in

A

miRNA Families	Start	End	Conservation Score			CLIP Support	Binding Sites in Structure
			Primates	Mammals	Vertebrates		
miR-30-5p	373	380	0.25	0.00	0.00		View
miR-129-5p	320	327	0.17	0.00	0.00		View
miR-130-3p/301-3p/454-3p	84	89	0.17	0.00	0.00		View
miR-139-5p	293	299	0.17	0.00	0.00		View
miR-142-5p	91	98	0.17	0.00	0.00		View
miR-146-5p	390	395	0.17	0.00	0.00		View
miR-21-5p/590-5p	206	211	0.17	0.00	0.00		View
miR-26-5p	339	345	0.17	0.00	0.00		View
miR-124-3p.1	221	226	0.08	0.00	0.00		View
miR-200bc-3p/429	187	192	0.08	0.00	0.00		View

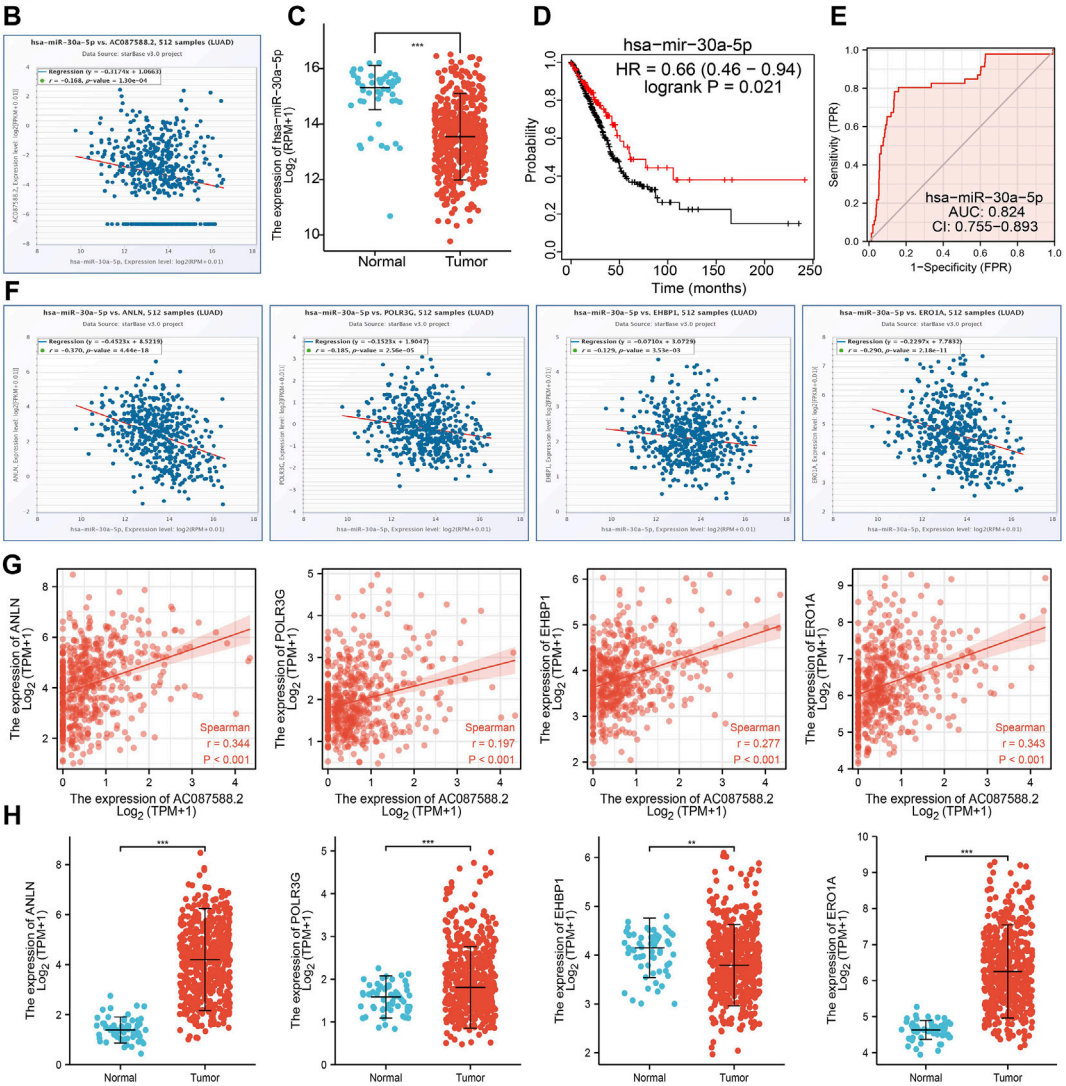


FIGURE 9 | Analysis of the potential miRNAs of AC087588.2. **(A)** Potential miRNAs of AC087588.2 determined by the Anbolnc2 database. **(B)**. Correlations between AC087588.2 expression and miR-30a-5p in LUAD. **(C)** Expression level of miR-30a-5p in LUAD. **(D)** Overall survival of miR-30a-5p in LUAD. **(E)** ROC curve of miR-30a-5p in LUAD. **(F)** Correlations between the miR-30a-5p expression and ANLN, POLR3G, EHBP1, and ERO1A in LUAD. **(G)** Correlations between the AC087588.2 expression and ANLN, POLR3G, EHBP1, and ERO1A in LUAD. **(H)** RNA level of ANLN, POLR3G, EHBP1, and ERO1A in LUAD.

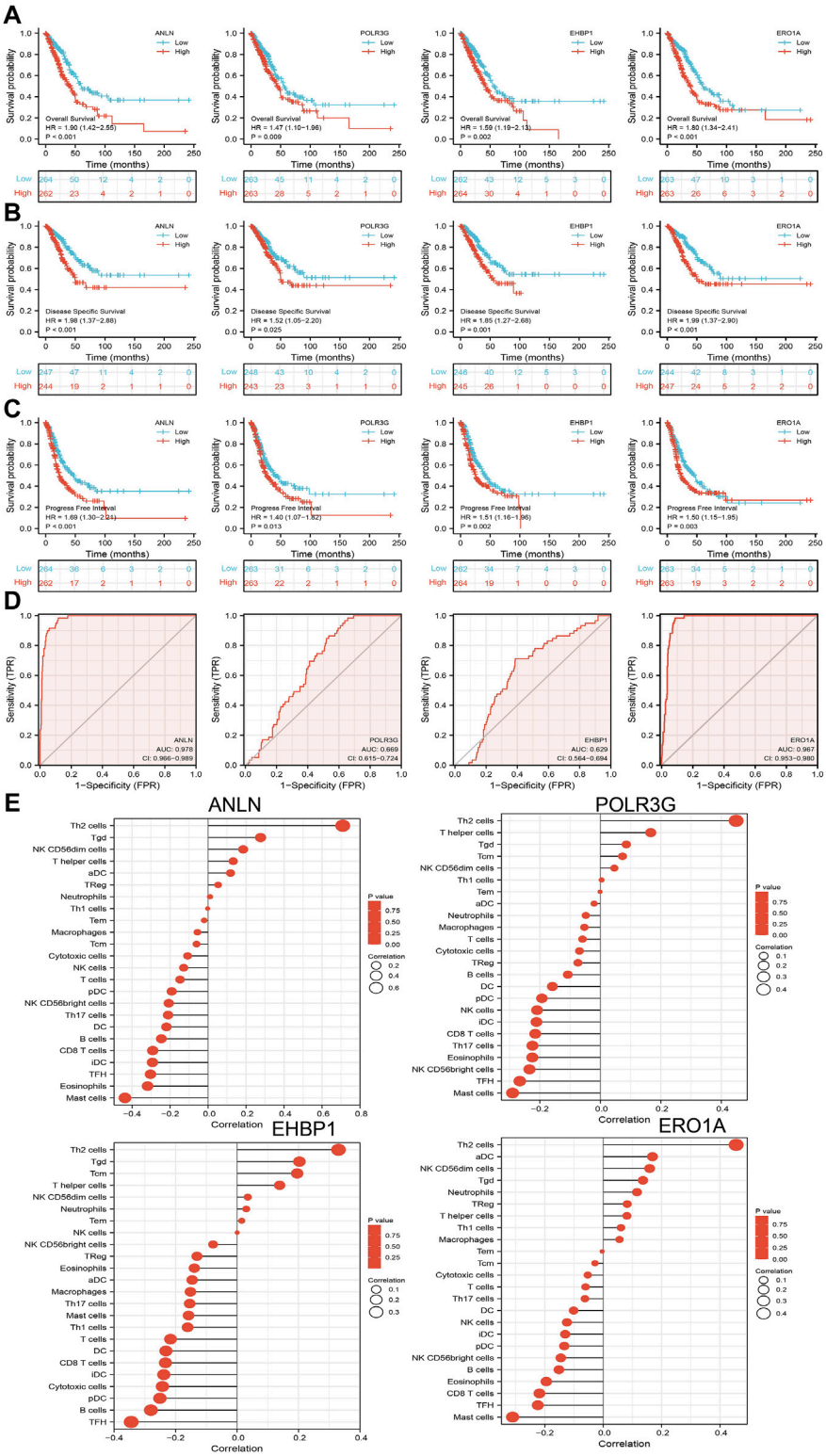


FIGURE 10 | Analysis of the prognostic and diagnostic value of ANLN, POLR3G, EHBP1, and ERO1A in LUAD. **(A–C)** OS, DSS, and PFS of ANLN, POLR3G, EHBP1, and ERO1A in LUAD. **(D)** ROC curve of ANLN, POLR3G, EHBP1, and ERO1A in LUAD. **(E)** Correlation between ANLN, POLR3G, EHBP1, and ERO1A expression and immune infiltration levels of 24 immune cells in LUAD.

LUAD. Given that immune checkpoints play a crucial role in tumor immunosuppression, we analyzed the correlation between the AC087588.2 expression and that of the immune checkpoint-related genes of *CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *PDCD1LG2*, *TIGIT*, and *SIGLEC15* in LUAD using Pearson's correlation analysis. The AC087588.2 expression was significantly positively correlated with the expression of *CD274*, *CTLA4*, *LAG3*, *TIGIT*, and *SIGLEC15* in this analysis (**Figure 8D**). These results confirmed that AC087588.2 played a crucial role in immune infiltration in LUAD.

Finally, we uncovered that AC087588.2 was significantly upregulated in NSCLC cell lines, and the depletion of AC087588.2 markedly inhibited cell proliferation and migration in LUAD.

We also utilized various databases to identify the potential target gene of AC087588.2/miRNA-30a-5p in LUAD, including the *ANLN*, *POLR3G*, *EHBP1*, and *ERO1A*. Subsequent expression and prognosis analysis confirmed that ANLN, POLR3G, EHBP1, and ERO1A were significantly greater in LUAD tissues than those in the normal LUAD tissues, and the upregulation of ANLN, POLR3G, EHBP1, and ERO1A expression was associated with poor prognosis in patients with LUAD. In conclusion, this finding provided a possible comprehensive view of the AC087588.2-mediated ceRNA network in LUAD, thereby highlighting its potential role in diagnosis and therapy. Finally, we uncovered that AC087588.2 was significantly upregulated in NSCLC cell lines and depletion of AC087588.2 markedly inhibited cell proliferation and migration in LUAD.

This study improves our understanding of the correlation between AC087588.2 and LUAD, but some limitations still exist. First, although we explored the correlation between AC087588.2 and immune infiltration in LUAD patients, there is a lack of experiments that validate the function of AC087588.2 in the tumor microenvironment regulation of LUAD. Second, we confirmed that the knockdown of AC087588.2 inhibited cell proliferation and cell migration of LUAD. However, the potential molecular mechanisms of AC087588.2 in cancer progression need to be explored in further studies.

CONCLUSION

Our data confirmed that AC087588.2 could serve as a promising diagnostic and prognostic biomarker for LUAD 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 authors.

AUTHOR CONTRIBUTIONS

XC, XJ, JG, and FZ designed this work and performed related assay. JP analyzed the data. LM and XN supervised and wrote the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Nature Science Foundation of China (82160512) and Yunnan Applied Basic Research Projects (2017FE467 and 2018FE001).

REFERENCES

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003

Chao, Y., and Zhou, D. (2019). lncRNA-D16366 Is a Potential Biomarker for Diagnosis and Prognosis of Hepatocellular Carcinoma. *Med. Sci. Monit.* 25, 6581–6586. doi:10.12659/msm.915100

Chen, Y., and Wang, X. (2020). miRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res.* 48 (D1), D127–d131. doi:10.1093/nar/gkz757

Cho, S., Jang, I., Jun, Y., Yoon, S., Ko, M., Kwon, Y., et al. (2013). MiRGator v3.0: a microRNA Portal for Deep Sequencing, Expression Profiling and mRNA Targeting. *Nucleic Acids Res.* 41, D252–D257. doi:10.1093/nar/gks1168

Deng, X., Xiong, W., Jiang, X., Zhang, S., Li, Z., Zhou, Y., et al. (2020). LncRNA LINC00472 Regulates Cell Stiffness and Inhibits the Migration and Invasion of Lung Adenocarcinoma by Binding to YBX1. *Cell Death Dis.* 11 (11), 945. doi:10.1038/s41419-020-03147-9

Dong, H. X., Wang, R., Jin, X. Y., Zeng, J., and Pan, J. (2018). LncRNA DGCR5 Promotes Lung Adenocarcinoma (LUAD) Progression via Inhibiting Hsa-mir-22-3p. *J. Cell Physiol.* 233 (5), 4126–4136. doi:10.1002/jcp.26215

Dweep, H., Gretz, N., and Sticht, C. (2014). miRWalk Database for miRNA-Target Interactions. *Methods Mol. Biol.* 1182, 289–305. doi:10.1007/978-1-4939-1062-5_25

Geng, Q., Li, Z., Li, X., Wu, Y., and Chen, N. (2021). LncRNA NORAD, Sponging miR-363-3p, Promotes Invasion and EMT by Upregulating PEAK1 and Activating the ERK Signaling Pathway in NSCLC Cells. *J. Bioenerg. Biomembr.* 53 (3), 321–332. doi:10.1007/s10863-021-09892-6

He, R., Zhang, F. h., and Shen, N. (2017). LncRNA FEZF1-AS1 Enhances Epithelial-Mesenchymal Transition (EMT) through Suppressing E-Cadherin and Regulating WNT Pathway in Non-small Cell Lung Cancer (NSCLC). *Biomed. Pharmacother.* 95, 331–338. doi:10.1016/j.biopha.2017.08.057

Jia, Y. C., Wang, J. Y., Liu, Y. Y., Li, B., Guo, H., and Zang, A. M. (2019). LncRNA MAFG-AS1 Facilitates the Migration and Invasion of NSCLC Cell via Sponging miR-339-5p from MMP15. *Cell Biol. Int.* 43 (4), 384–393. doi:10.1002/cbin.11092

Ke, L., Yang, D.-C., Wang, Y., Ding, Y., and Gao, G. (2020). AnnoLnc2: the One-Stop Portal to Systematically Annotate Novel lncRNAs for Human and Mouse. *Nucleic Acids Res.* 48 (W1), W230–w238. doi:10.1093/nar/gkaa368

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks from Large-Scale CLIP-Seq Data. *Nucleic Acids Res.* 42, D92–D97. doi:10.1093/nar/gkt1248

Li, T., Xie, J., Shen, C., Cheng, D., Shi, Y., Wu, Z., et al. (2016). Upregulation of Long Noncoding RNA ZEB1-AS1 Promotes Tumor Metastasis and Predicts Poor Prognosis in Hepatocellular Carcinoma. *Oncogene* 35 (12), 1575–1584. doi:10.1038/onc.2015.223

Lu, W., Zhang, H., Niu, Y., Wu, Y., Sun, W., Li, H., et al. (2017). Long Non-coding RNA Linc00673 Regulated Non-small Cell Lung Cancer Proliferation,

Migration, Invasion and Epithelial Mesenchymal Transition by Sponging miR-150-5p. *Mol. Cancer* 16 (1), 118. doi:10.1186/s12943-017-0685-9

Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., and Adjei, A. A. (2008). Non-small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clin. Proc.* 83 (5), 584–594. doi:10.4065/83.5.584

Pan, J., Fang, S., Tian, H., Zhou, C., Zhao, X., Tian, H., et al. (2020). lncRNA JPX/miR-33a-5p/Twist1 axis Regulates Tumorigenesis and Metastasis of Lung Cancer by Activating Wnt/β-Catenin Signaling. *Mol. Cancer* 19 (1), 9. doi:10.1186/s12943-020-1133-9

Qi, Q., Chen, C., Liu, C., Zhang, B., Ma, Y., Zhang, H., et al. (2021). Linc8087 Predicts Favorable Prognosis and Inhibits Cell Migration and Invasion in NSCLC. *Pathology - Res. Pract.* 225, 153569. doi:10.1016/j.prp.2021.153569

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an Immeasurable Source of Knowledge. *Contemp. Oncol. Pozn.* 19 (1a), A68–A77. doi:10.5114/wo.2014.47136

Vasaikar, S. V., Straub, P., Wang, J., and Zhang, B. (2018). LinkedOmics: Analyzing Multi-Omics Data within and across 32 Cancer Types. *Nucleic Acids Res.* 46 (D1), D956–d963. doi:10.1093/nar/gkx1090

Wang, L.-F., Wu, L.-P., and Wen, J.-D. (2021). LncRNA AC079630.4 Expression Associated with the Progression and Prognosis in Lung Cancer. *Aging* 13 (14), 18658–18668. doi:10.18632/aging.203310

Xue, L., Li, J., Lin, Y., Liu, D., Yang, Q., Jian, J., et al. (2021). A Transferase METTL3-Induced lncRNA ABHD11-AS1 Promotes the Warburg Effect of Non-small-cell Lung Cancer. *J. Cell Physiol.* 236 (4), 2649–2658. doi:10.1002/jcp.30023

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhang, M., Wang, N., Song, P., Fu, Y., Ren, Y., Li, Z., et al. (2020). LncRNA GATA3-AS1 Facilitates Tumour Progression and Immune Escape in Triple-Negative Breast Cancer through Destabilization of GATA3 but Stabilization of PD-L1. *Cell Prolif.* 53 (9), e12855. doi:10.1111/cpr.12855

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 Jiang, Chen, Guo, Zhou, Pu, Mutti and Niu. 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 Physiology

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fphys.2022.905523

ESRRG, ATP4A, and ATP4B as Diagnostic Biomarkers for Gastric Cancer: A Bioinformatic Analysis Based on Machine Learning

Qiu Chen[1†], *Yu Wang*[2†], *Yongjun Liu*[2] *and Bin Xi*[2*]

[1]*Medical College, Yangzhou University, Yangzhou, China,* [2]*College of Physics Science and Technology, Yangzhou University, Yangzhou, China*

OPEN ACCESS

Edited by:
William C. Cho,
QEH, Hong Kong SAR, China

Reviewed by:
Yuan Zhu,
China University of Geosciences Wuhan, China
Swakkhar Shatabda,
United International University, Bangladesh

***Correspondence:**
Bin Xi
xibin@yzu.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

Received: *27 March 2022*
Accepted: *10 May 2022*
Published: *23 June 2022*

Citation:
Chen Q, Wang Y, Liu Y and Xi B (2022) ESRRG, ATP4A, and ATP4B as Diagnostic Biomarkers for Gastric Cancer: A Bioinformatic Analysis Based on Machine Learning. Front. Physiol. 13:905523. doi: 10.3389/fphys.2022.905523

Based on multiple bioinformatics methods and machine learning techniques, this study was designed to explore potential hub genes of gastric cancer with a diagnostic value. The novel biomarkers were detected through multiple databases of gastric cancer–related genes. The NCBI Gene Expression Omnibus (GEO) database was used to obtain gene expression files. Three hub genes (*ESRRG*, *ATP4A*, and *ATP4B*) were detected through a combination of weighted gene co-expression network analysis (WGCNA), gene–gene interaction network analysis, and supervised feature selection method. GEPIA2 was used to verify the differences in the expression levels of the hub genes in normal and cancer tissues in the RNA-seq levels of Genotype-Tissue Expression (GTEx) and The Cancer Genome Atlas (TCGA) databases. The objectivity of potential hub genes was also verified by immunohistochemistry in the Human Protein Atlas (HPA) database and transcription factor–hub gene regulatory network. Machine learning (ML) methods including data pre-processing, model selection and cross-validation, and performance evaluation were examined on the hub-gene expression profiles in five Gene Expression Omnibus datasets and verified on a GEO external validation (EV) dataset. Six supervised learning models (support vector machine, random forest, k-nearest neighbors, neural network, decision tree, and eXtreme Gradient Boosting) and one semi-supervised learning model (label spreading) were established to evaluate the diagnostic value of biomarkers. Among the six supervised models, the support vector machine (SVM) algorithm was the most effective one according to calculated performance metrics, including 0.93 and 0.99 area under the curve (AUC) scores on the test and external validation datasets, respectively. Furthermore, the semi-supervised model could also successfully learn and predict sample types, achieving a 0.986 AUC score on the EV dataset, even when 10% samples in the five GEO datasets were labeled. In conclusion, three hub genes (*ATP4A*, *ATP4B*, and *ESRRG*) closely related to gastric cancer were mined, based on which the ML diagnostic model of gastric cancer was conducted.

Keywords: gastric cancer, machine learning, bioinformatics, WGCNA, diagnostic model

1 INTRODUCTION

Gastric cancer (GC), reported as the sixth most common cancer in the world, has an extremely high morbidity rate (Sung et al., 2021). Latest global epidemiological data showed that almost 1,089,103 people were diagnosed with gastric cancer every year, and 768,793 people died of this disease, which makes it the fourth most fatal cancer worldwide (Sung et al., 2021). Although previous research studies have successfully revealed the major risk factors of GC, such as the genetic background, obesity, harmful mode of life, and *Helicobacter pylori* infection, a high rate of misdiagnosis still exists due to nonspecific symptoms at the beginning of the disease (Van Cutsem et al., 2016). In other words, GC usually has a late diagnosis at an advanced stage, resulting in its proximity to morbidity and mortality (Asplund et al., 2018). The prognosis of locally advanced gastric cancer is poor with a 5-year survival rate of 16.4% (Katai et al., 2018) and median overall survival (OS) of 6–14 months in East Asia after being diagnosed from extensive clinical studies (Hu et al., 2021). In contrast, if GC is diagnosed at an early stage, the 5-year survival rate is about 90% (Saragoni et al., 2013), indicating the importance of early diagnosis and treatment. Novel biomarkers screened through bioinformatics methods have already shown their potentiality in cancer development and diagnosis. Therefore, it is extremely meaningful to find novel biomarkers of GC to assist in the early diagnosis and treatment.

Recently, machine learning (ML) has been widely used as a bioinformatics method in the realm of medical data mining (Yang et al., 2020). Compared with traditional analyses, the ML technique has an edge on discovering hidden relationships and making predictions from complex datasets which have already been successful in many clinical practices, such as image-based cancer screening (Hu et al., 2018), constructing effective prognostic models (Royston et al., 2004), and identifying biomarkers based on the integration of omics and phenotype data (Subramanian et al., 2020). On the other side, biological networks such as weighted co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) and gene–gene interaction networks can identify the associations between genes and the biological processes. In accordance with biological network analyses, novel genes and pathways related to human cancers are also revealed (Boucher and Jenna, 2013; Farhadian et al., 2021). Thus, combing the core concepts of ML such as feature selection and classification with additional biological network analyses may further assist in exploring biomarkers with diagnostic values.

In this study, our purpose was to explore biomarkers based on biological network analyses and ML techniques, the novelty of which is further examined with ML diagnostic models. Potential hub genes are screened by the feature selection method and biological networks. ML diagnostic models are constructed by supervised and semi-supervised ML methods with stratified k-fold cross-validation and random permutation validation, respectively.

2 MATERIALS AND METHODS

2.1 Data Collection and Preprocessing

The study design is shown in **Figure 1**. This systematic study comprehensively downloaded six datasets from the Gene Expression Omnibus (GEO) database and focused on the gene sequencing results of GC patients with each dataset containing more than 10 samples. These datasets were produced using three different microarray platforms: Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Human Exon 1.0 ST Array, and Affymetrix Human Genome U133A Array. Raw data of these datasets were preprocessed by R packages "oligo" (Carvalho and Irizarry, 2010) and "affy" (Gautier et al., 2004), and then, the background was corrected and normalized through the Robust Multichip Average (RMA) function. In this study, GSE66229 was used to construct a weighted gene co-expression network due to the sufficient data and detailed clinical characteristics of the gastric cancer samples. Five datasets (GSE19826, GSE27342, GSE29272, GSE54129, and GSE66229) were combined into a total dataset (TD) which contains 780 samples and 11,181 genes for feature selection and building ML models. TD includes 435 tumor samples and 345 normal ones, i.e., a mild imbalanced dataset. The combat algorithm in the "sva" R package (Johnson et al., 2007) was used to eliminate batch effects between different platforms and experiments. GSE33335 acts as an independent dataset, based on which an external validation (EV) was performed to validate the authenticity of hub genes and the reproducibility and generalizability of the ML diagnostic models. Details of all datasets can be found in **Supplementary Table S1**.

2.2 WGCNA

The R package "WGCNA" (Langfelder and Horvath, 2008) was constructed to detect gene modules, and the correlation of each module with sample type was evaluated. The specific steps are as follows: (a) in the GSE66229 dataset, only normal and cancer samples from the same individuals (196 samples) were selected for further analysis. Then, the 196 samples were divided into "tumor" and "normal" groups according to their clinical records, with each group containing 98 samples; (b) the samples were clustered by the "hclust" function to detect the outliers. After employing the "hclust" function to the expression matrix evaluated by the average method, 35 offending samples were removed with a height cut at 125; (c) the best scale-free topology fitting index (soft threshold) was selected as 7 to achieve a higher average network connectivity with a scale-free fitting number $\beta = 0.86$; d) the adjacency matrix was transformed into a topological overlap matrix (TOM) to define the gene co-expression similarity; (e) Based on the dissimilarity measured by TOM, the "hclust" algorithm was employed for gene hierarchical clustering; (f) the optimal module size was set as 30, and the dynamic tree was used to cut the identification module; (g) after each module was determined based on the signature gene expression profile and the

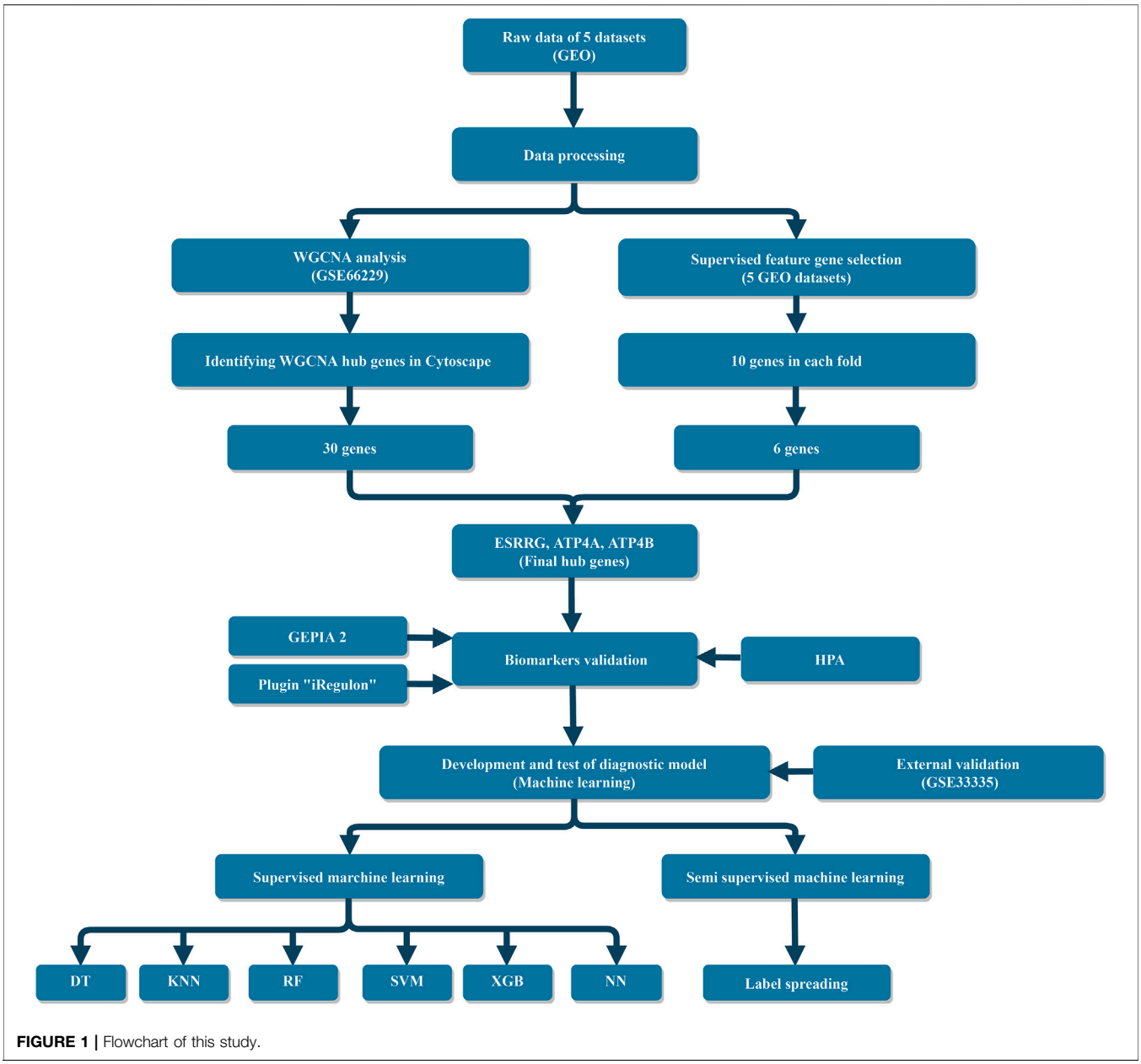


FIGURE 1 | Flowchart of this study.

sample type of patients, the correlation of the module signature genes with sample types was also determined.

2.3 Identification of the WGCNA Hub Genes

Cytoscape (Shannon et al., 2003) was used to visualize the co-expression network in the modules of the highest correlations. All genes in the selected modules were exported to Cytoscape and analyzed with the "NetworkAnalyzer" plugin (Assenov et al., 2007), which can give a comprehensive set of topological parameters for gene–gene interaction networks. Hub genes are defined as genes with high connectivity in the gene–gene interaction network. According to connectivity, i.e., node degrees in the output of "NetworkAnalyzer", the top-ranked 10% genes in the two most significant modules "red" and "turquoise" were selected (Fuxman Bass et al., 2013), which may have important implications for the progression of gastric cancer.

2.4 Supervised Feature Gene Selection With the Fisher Score Algorithm

The feature selection technique is a process of reducing the number of variables, especially important for developing a predictive model (Ali et al., 2018). The feature selection method can evaluate the relationship between each variable and the output and select those variables with the strongest relationship. Fisher score is one of the most widely used supervised feature selection methods, which returns the ranks

TABLE 1 | All genes and their Fisher Scores were selected by the feature algorithm.

Fold 1		Fold 2		Fold 3		Fold 4		Fold 5	
Gene name	**Fisher score**	**Gene name**	**Fisher score**	**Gene name**	**Fisher score**	**Gene name**	**Fisher score**	**Gene name**	**Fisher score**
ATP4A	0.762	ATP4A	0.745	ATP4A	0.796	ATP4A	0.862	ATP4A	0.810
ESRRG	0.736	ESRRG	0.705	ESRRG	0.734	ESRRG	0.803	ESRRG	0.749
CBLIF	0.671	ATP4B	0.642	CBLIF	0.642	CBLIF	0.748	CBLIF	0.670
ATP4B	0.641	CBLIF	0.632	ATP4B	0.631	ATP4B	0.712	ATP4B	0.644
INHBA	0.548	TIMP1	0.574	SST	0.540	INHBA	0.618	KCNE2	0.553
KCNE2	0.541	KCNE2	0.517	MT1M	0.539	KCNE2	0.615	TIMP1	0.541
CPA2	0.533	INHBA	0.513	TIMP1	0.538	CPA2	0.601	INHBA	0.524
MT1M	0.531	CPA2	0.498	INHBA	0.522	TIMP1	0.595	CPA2	0.520
ALDH6A1	0.529	MT1M	0.491	KCNE2	0.511	MYRIP	0.587	MYRIP	0.505
TIMP1	0.510	GKN1	0.468	GKN1	0.501	MT1M	0.543	MT1M	0.494

of the variables based on the Fisher score in the descending order (Gu et al., 2011). The Fisher score S_i of the i-th feature is calculated as follows:

$$S_i = \frac{\sum_j n_j \left(\mu_{ij} - \mu_i\right)^2}{\sum_j n_j \sigma_{ij}^2}, \quad (1)$$

where μ_{ij} and σ_{ij} are the mean and standard deviation of the i-th feature in the j-th class, respectively. n_j is the number of samples in the j-th class, and μ_i is the mean of the i-th feature.

In this study, to select the most relevant genes that are strongly related to the sample type, feature selection using the Fisher score algorithm was applied to the combined five datasets. Here, a gene was regarded as a feature, and TD was splitted into five folds during feature selection. A list of genes ranked by their scores returned in each fold, where we picked the top-ranked 10 feature genes with a cutoff at $S_i \approx 0.5$ for each list for further study. The feature genes were determined as the intersection of the features of five folds. The final biomarkers in this study were obtained by the intersection of the hub genes filtered by the gene–gene interaction network and the feature genes.

2.5 Validation of the Final Hub Genes

GEPIA2 can be used to verify the expression difference of the hub genes in tumor samples and normal ones (Tang et al., 2019). The RNA-seq datasets used in GEPIA2 were based on UCSC Xena (http://xena.ucsc.edu), which was computed by standard pipelines to analyze the RNA-sequencing expression of tumor and normal samples from the TCGA (Colaprico et al., 2016) and GTEx (Lonsdale et al., 2013) datasets. In this study, we used the TCGA and GTEx gastric cancer RNA-seq data integrated by the GEPIA2 platform for a comprehensive validation. With $|\mathrm{Log_2FC}|$ cutoff = 1 and p-value cutoff = 0.01, box plots of the RNA-seq data of the gastric cancer hub genes were drawn.

The immunohistochemistry (IHC) staining data for this study were downloaded from the Human Protein Atlas (HPA) database (Thul and Lindskog, 2018), and then, the results of gastric cancer pathology and normal gastric tissue were processed.

The Cytoscape plugin "iRegulon" was used to analyze the transcription factors regulating hub genes (Janky et al., 2014; Gao et al., 2020). This plugin predicts transcription factors by using the motif enrichment analysis and using track discovery in a set of regulated genes. The cutoff criteria were as follows: enrichment score threshold = 3.0, receiver operating characteristic (ROC) threshold for area under the curve (AUC) calculation = 0.03, rank threshold = 5,000, minimum identity between orthologous genes = 0.0, and false discovery rate (FDR) = 0.001. After all transcription factors were outputted, the factor which regulates all hub genes and ranks first in the normalized enrichment score (NES) was defined as the most relevant transcription factor.

2.6 Development and Validation of Machine Learning Models

2.6.1 Supervised Learning

At the first step, TD was randomly split into training and test datasets, with ratios of 80 and 20%, respectively. Then, a repeated stratified k-fold cross-validation was performed on the training dataset. The stratified k-fold can ensure that each fold has the same proportion of the sample type compared to the whole one, which is more suitable to imbalance datasets. The ML model was trained using of k-1 folds and validated on the one remaining fold for k times. The training performance of the model was reported on the average over k times. At last, a final evaluation was performed on the test dataset. The aforementioned steps can be regarded as an internal validation since both training and test datasets come from TD. To examine the robustness of the ML models, an EV was further performed on the independent dataset GSE33335.

In this study, k was set to 10, and the cross-validation was repeated 100 times with different randomizations in each repetition to ensure the estimated performance. The Matthews correlation coefficient (MCC) metric (Chicco and Jurman, 2020) was chosen as the performance score for the model evaluation during the training process, which is suitable to imbalance datasets.

To further reduce the affection of the dataset imbalance, the synthetic minority oversampling technique (SMOTE) was applied to the training dataset (Chawla et al., 2002). The SMOTE can synthesize new samples based on randomly picked existing samples and their k-nearest neighbors. In this study, a grid search of k ranging from 1 to 7 was also performed.

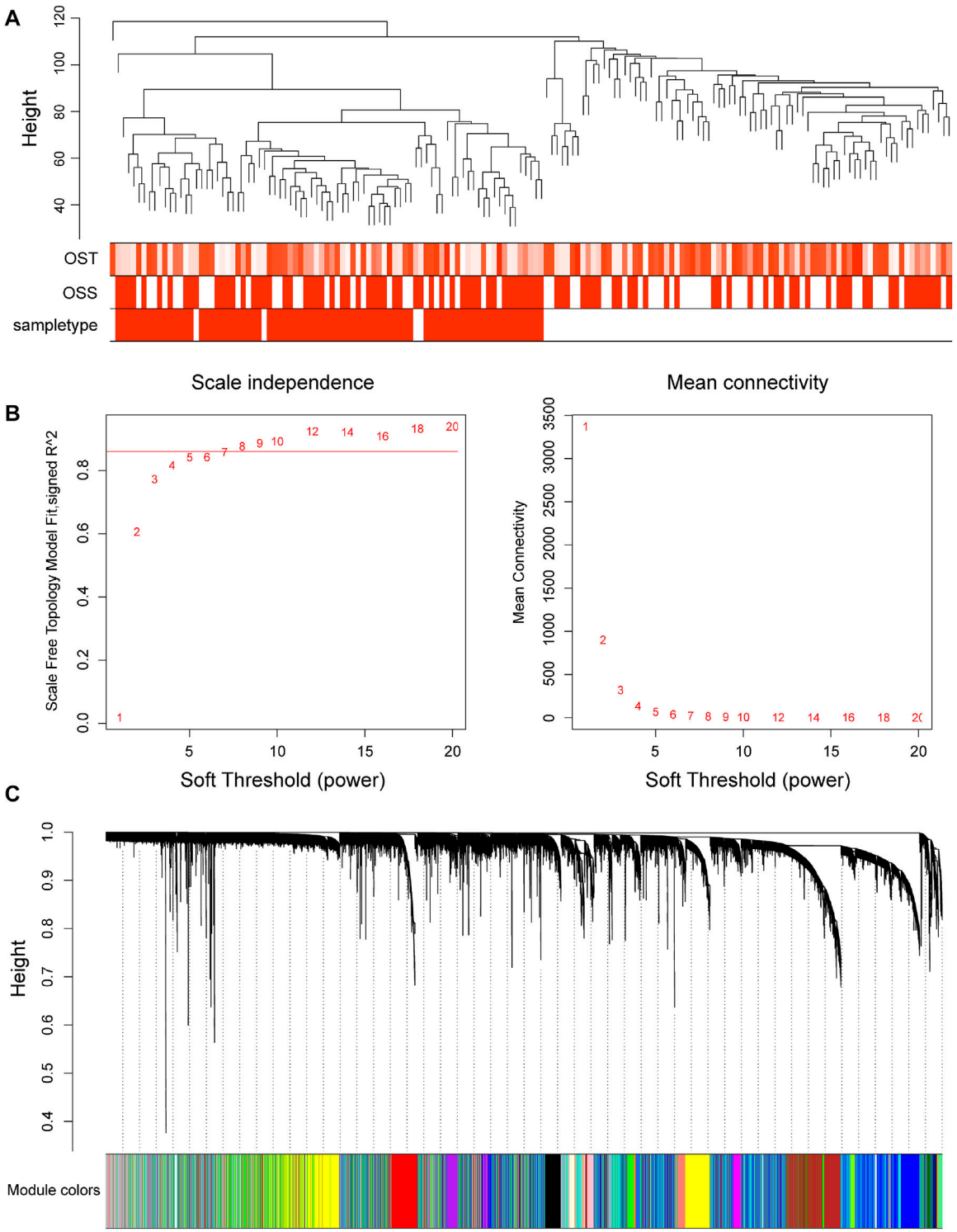


FIGURE 2 | Progress of the weighted gene co-expression network analysis in GSE66229. **(A)** Cluster dendrogram of 161 samples in GSE66229. **(B)** Soft thresholds of the best scale-free topological model fitting index (left) and mean connectivity (right) were determined. The red horizontal line represents $R^2 = 0.86$. **(C)** Dendrogram of all genes clustered in GSE66229. Gene clustering into modules is based on a topological overlap matrix. Assigned modules are colored on the bottom with gray denoting unassigned genes.

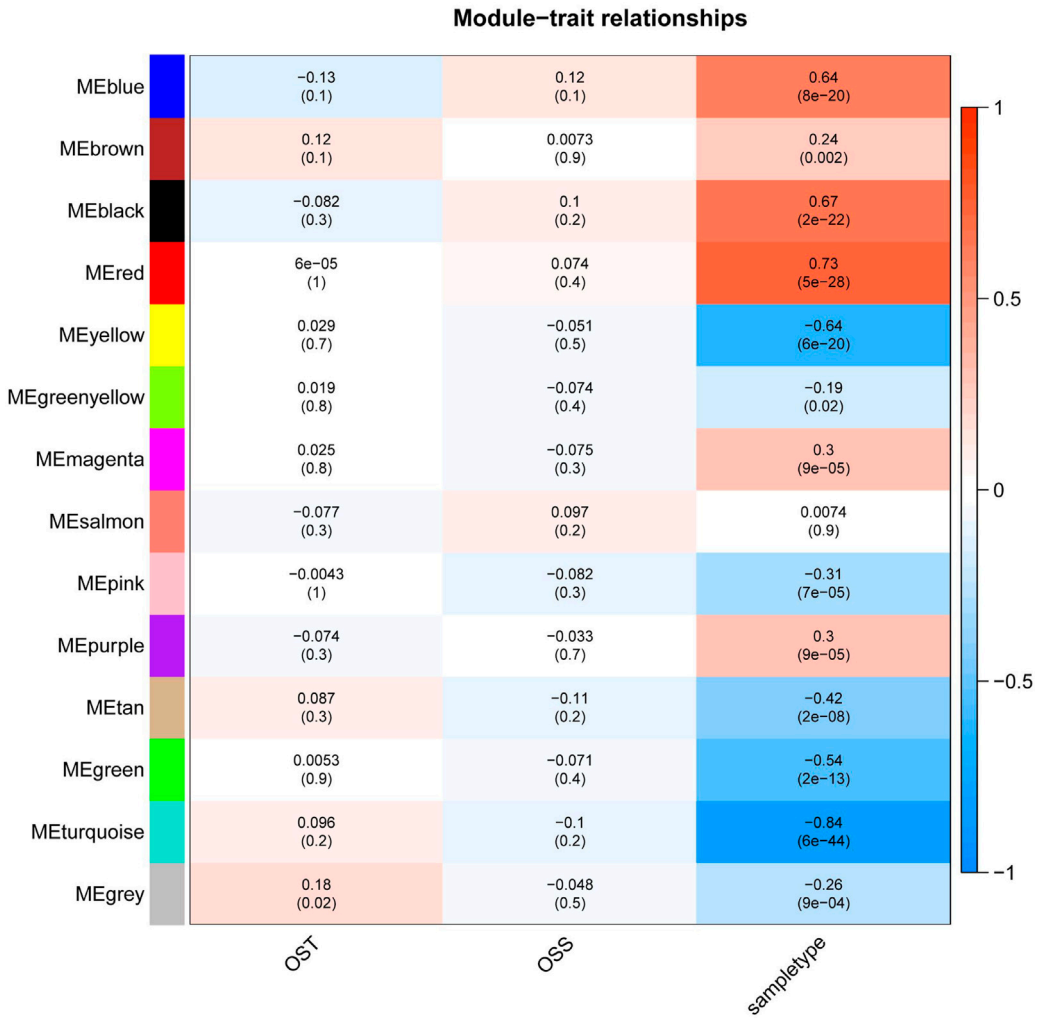


FIGURE 3 | Heatmap of the relationship between module eigengenes and clinical traits of GSE66229. WGCNA labeled heatmaps for GSE66229, each row represents a module characteristic gene encoded by color, and the three columns represent clinical characteristics of overall survival time (OST), overall survival status (OSS), and sample type, respectively. Each cell represents the Pearson correlation coefficient and *p*-value (in parentheses) of the corresponding module characteristics, and the color of each cell represents the value of correlation.

In order to select a proper classifier for the ML diagnosis model, six widely used algorithms, namely, support vector machine (SVM) (Byvatov and Schneider, 2003), k-nearest neighbors (KNN) (Zhang, 2016), decision tree (DT) (Chen et al., 2011), random forest (RF) (Chen and Ishwaran, 2012), neural network (NN) (Lancashire et al., 2009), and eXtreme Gradient Boosting (XGB) (Chen and Guestrin, 2016), were examined through their performance metrics for classification results. Hyperparameters of all the models were finely tuned using the scikit-learn GridSearchCV method, according to the highest "MCC" scores. The best model for each algorithm was selected after exploration of the whole grid. Best hyperparameters and the corresponding training performances of all supervised ML diagnostic models can be found in **Table 1**. Finally, the performance of each model on test and EV datasets was evaluated by these performance metrics: accuracy (Heidaryan, 2018), specificity (Altman and Bland, 1994), sensitivity (Altman and Bland, 1994), precision (Heidaryan, 2018), F1 score (Chicco and Jurman, 2020), and MCC. Furthermore, the ROC curve and AUC are also given.

2.6.2 Semi-Supervised Learning

To deal with a different problem, such as handling large amounts of samples with only a few diagnosed ones, a semi-supervised learning model based on a label-spreading algorithm is also examined (Zhou et al., 2003). Semi-supervised learning can learn from small amounts of labeled samples, combined with the use of unlabeled data to better capture the underlying properties and generalize better to new samples (Chapelle et al., 2009). To some extent, semi-supervised learning can be regarded as a hybrid of supervised and unsupervised learning. In this study, TD was randomly split into a labeled dataset and an unlabeled one, with five unlabeled ratios including 50, 60, 70, 80,

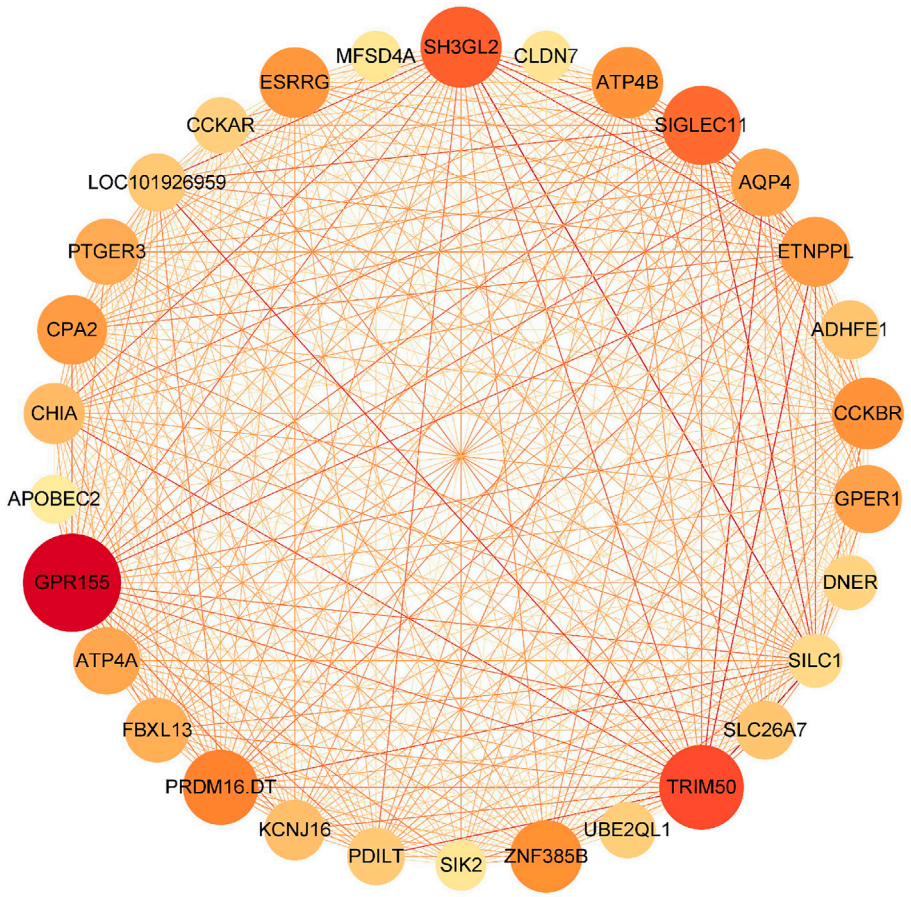


FIGURE 4 | Gene–gene interaction network of the top-ranked 10% genes in red modules.

and 90%. For each ratio, the semi-supervised model was cross-validated by 100 times of random permutations and further evaluated on the EV dataset GSE33335. The performance metrics of prediction on the unlabeled dataset and EV dataset are given.

All supervised and semi-supervised ML models in this study were implemented by Python language programming on Intel Xeon silver 4110 CPU.

3 RESULTS

3.1 Construction of the Gene Co-Expression Network

In order to find the correlation between clinical features and genes, this study used the R "WGCNA" package to construct 20,862 genes and 161 samples in the GSE66229 dataset into a gene network. A sample clustering figure was plotted (**Figure 2A**). To guarantee a scale-free topology and zero mean connectivity, the threshold was determined to be 7 (**Figure 2B**). The dissimilarity of the modules was set as 0.2, and a total of 14 modules were generated (**Figure 2C**). Two modules (red: r = 0.73 and P = 5e-28; turquoise: r = -0.84 and P = 6e-44) with the positive and negative highest correlations were acquired as the significant modules for subsequent analyses (**Figure 3**).

3.2 Feature Gene Selection

In order to select critical genes to the diagnostic model, feature selection was performed for the combined five datasets using the Fisher Score method on five folds. In each fold, a cutoff around the Fisher Score $S_i \approx 0.5$ s was applied, and as a result, 10 genes with the highest scores were selected. All selected genes as well as their Fisher Scores are listed in **Table 1**. At last, the intersection of all picked genes in the five folds is investigated, resulting in six intersection elements: TIMP1, ATP4A, ESRRG, CBLIF, ATP4B, and INHB.

3.3 Identification and Validation of Hub Genes

In the results of WGCNA, two significant models, the red and turquoise ones, were exported to Cytoscape. Two gene–gene interaction networks were constructed and analyzed in Cytoscape. Then, the top 10% target genes of each network were selected, according to the connectivity degree. As a

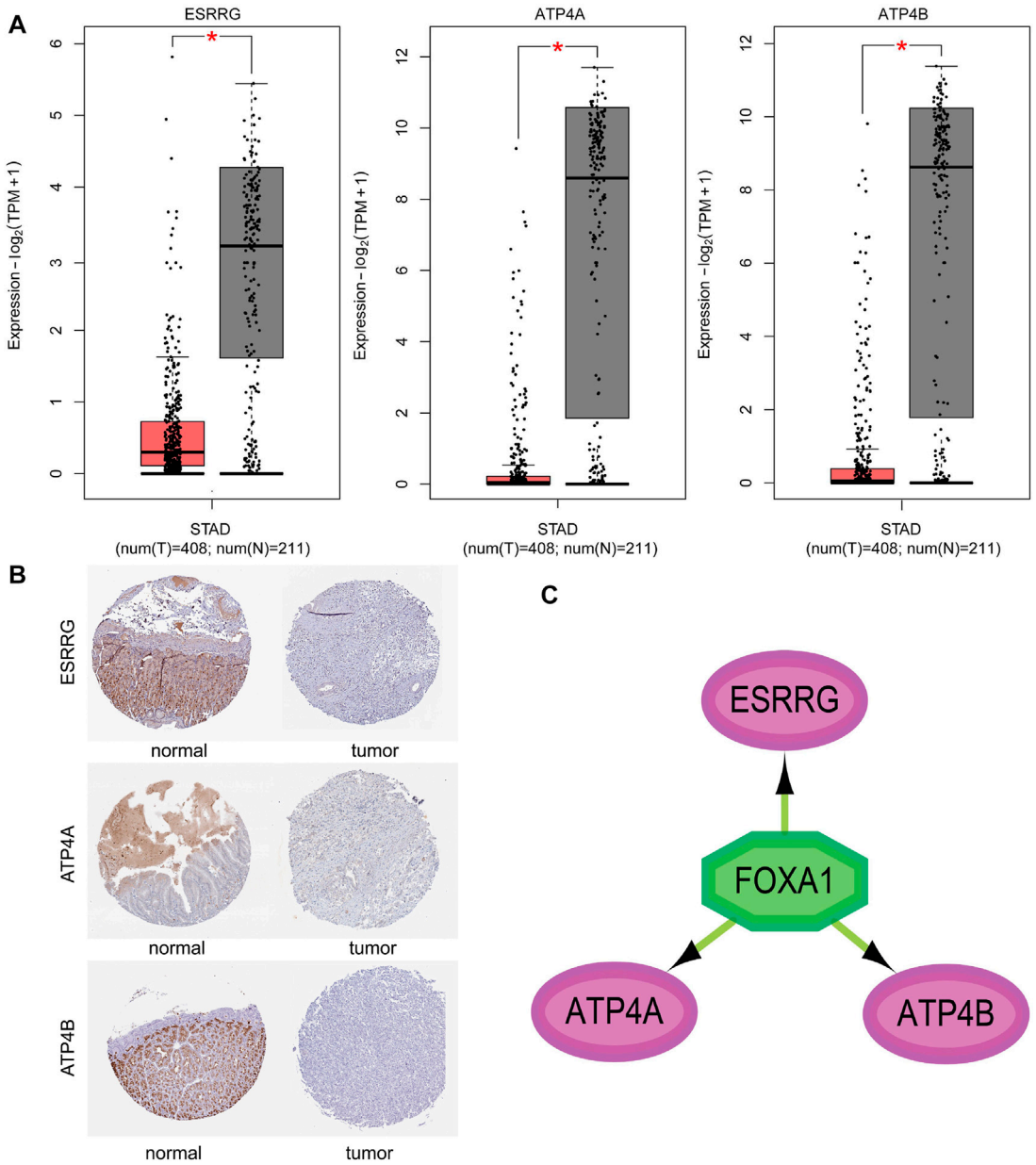


FIGURE 5 | Validation of three hub gene expressions in the GEPIA2 platform. **(A)** Validation of three hub gene expressions in the GEPIA2 platform. The red and gray boxes represent cancer and normal tissues in the TCGA and GTEx datasets, respectively. STAD, gastic cancer, and $p < 0.01$ (GEPIA2 website). **(B)** Immunohistochemical staining of ESRRG, ATP4A, and ATP4B in the Human Protein Atlas (HPA) database. **(C)** Transcription factor–hub gene regulatory network of the most relevant factor in the Cytoscape plugin "iRegulon".

TABLE 2 | Tuned hyperparameters, k in the SMOTE, and the training performance of six machine learning models.

Model	Tuned hyperparameters	k	MCC
SVM	C: 3,000 and gamma: 0.01	5	0.666 ± 0.093
RF	max_features: log2 and n_estimators: 1,500	5	0.641 ± 0.091
KNN	metric: manhattan and n_neighbors: 29	3	0.649 ± 0.089
NN	activation: tanh and hidden_layer_sizes: (200, 200, and 200)	6	0.633 ± 0.096
DT	max_depth: 60, min_impurity_decrease: 0.2, and min_samples_leaf: 2	7	0.637 ± 0.090
XGB	gamma: 1, max_depth: 2, and n_estimators: 100	5	0.658 ± 0.092

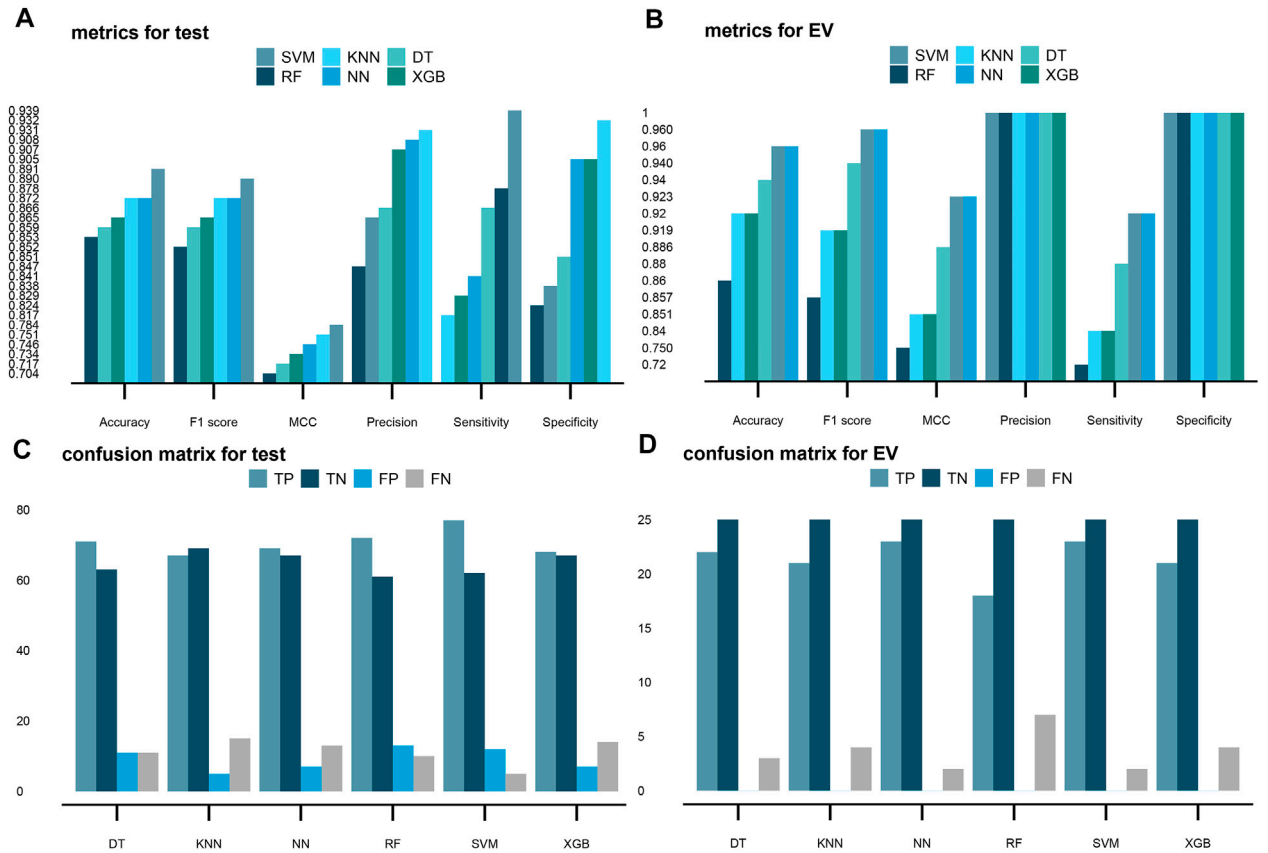


FIGURE 6 | Performance of the six supervised machine learning models on the test and EV sets. Hyperparameters of all six models are tuned with the GridSearchCV method, according to the "MCC" metric, and then, the six best models were chosen after exploration of the whole grid. Predictions on the test and EV sets are made with the best models. Six models used in this study are support vector machine (SVM), k-nearest neighbors (KNN), decision tree (DT), random forest (RF), neural network (NN), and eXtreme Gradient Boosting (XGB) in order. **(A,B)** Scores of accuracy, F1 score, MCC, precision, sensitivity, and specificity in the six models on the test and valid datasets, respectively. **(C,D)** Four terms of the confusion matrix (TP, TN, FP, and FN) in the six models on the test and valid datasets, respectively.

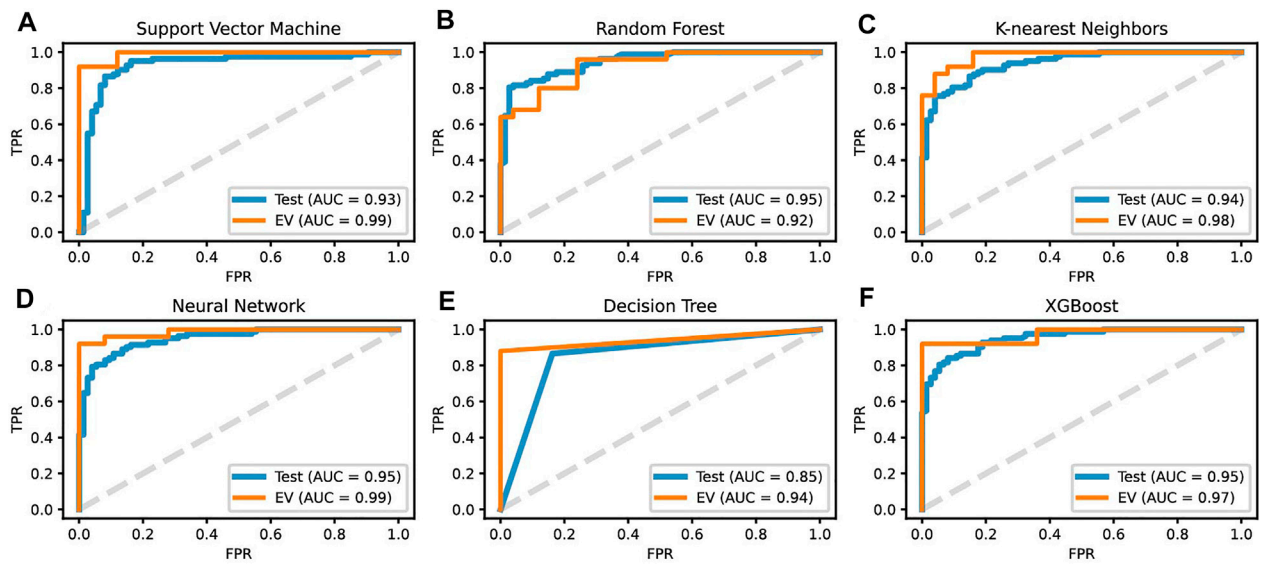


FIGURE 7 | ROC curves for the predicted probability on the test and EV sets of all six machine learning diagnostic models: **(A)** SVM, **(B)** RF, **(C)** KNN, **(D)** NN, **(E)** DT and **(F)** XGB.

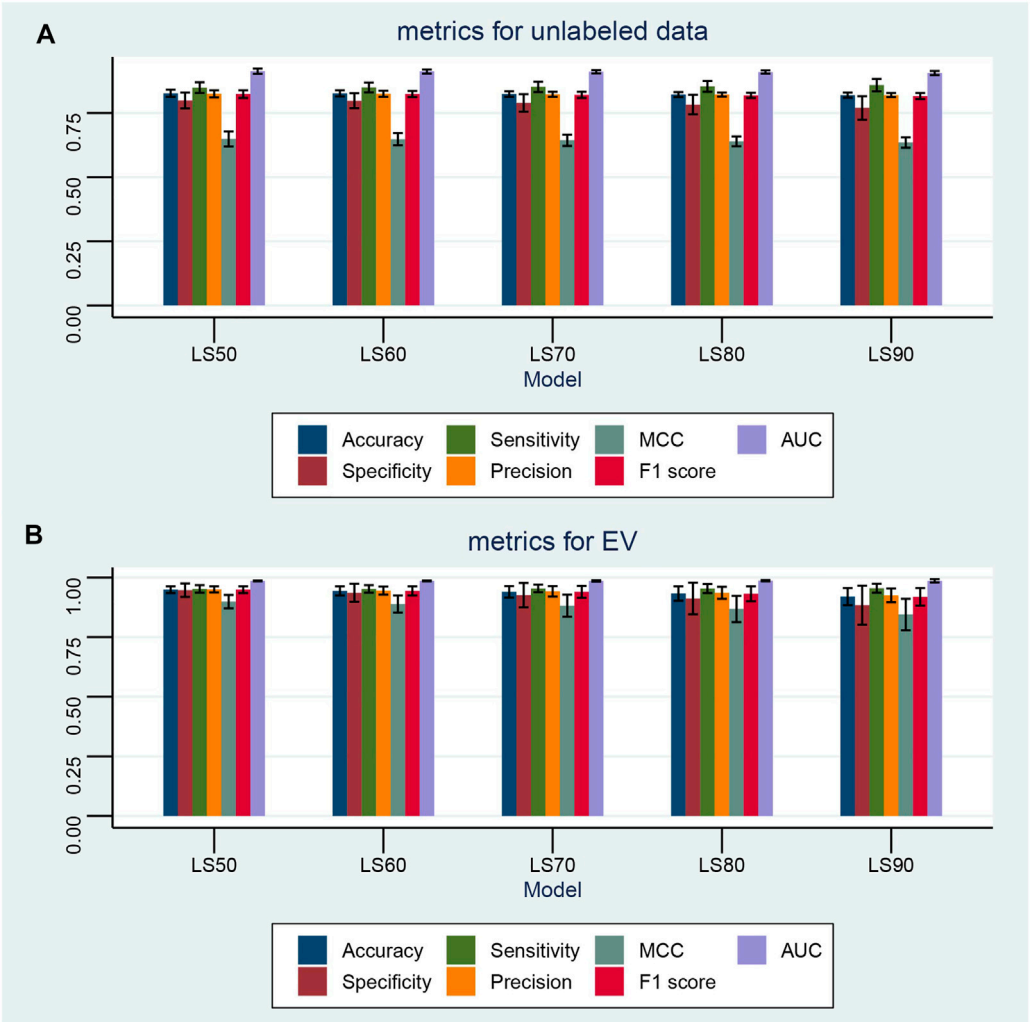


FIGURE 8 | Performance of the semi-supervised machine learning model with various ratios of unlabeled data. Semi-supervised machine learning models are built with the label spreading (LS) algorithm. The ratios of randomly unlabeled samples include 50% (LS50), 60% (LS60), 70% (LS70), 80% (LS80), and 90% (LS90). In each ratio, the semi-supervised model is cross-validated 100 times by random permutation. **(A,B)** Performance of the semi-supervised machine learning models on all unlabeled data and the valid dataset with various ratios of unknown samples, respectively. Seven metrics are given, namely, accuracy, F1 score, MCC, precision, sensitivity, specificity, and AUC.

result, 30 and 330 genes in the "red" and "turquoise" modules were selected. The gene–gene network of 30 genes in the "red" module is shown in (**Figure 4**), while genes in the "turquoise" module are listed in **Supplementary material S1**. Together with the six feature genes selected through the Fisher Score method, three hub genes (*ESRRG*, *ATP4A*, and *ATP4B*) were finally selected in the red module, while none was selected in the turquoise module.

The expression of three genes in cancer and normal samples was validated in GEPIA2. The box plot of GEPIA2 presented the expression levels of the three genes in the standard of expression-log2 (TPM+1) (**Figure 5A**). We observed that the expressions of ESSRG, ATP4A, and ATP4B in tumor samples were significantly lower than those in normal ones. This study also performed an IHC analysis in the gastric data of hub genes from the HPA database. The results of IHC staining are shown in **Figure 5B**, which were consistent with GEPIA2.

Among all transcription factors regulating the three hub genes, FOXA1 with the highest NES (NES = 5.142) was considered as the most important transcription factors (**Figure 5C**). Existing studies have shown that the expression of FOXA1 affects the proliferation and invasion of gastric cancer cells (Lin et al., 2018; Dai Y. et al., 2021). The result verified the objectivity of the three hub genes in gastric cancer.

3.4 Establishment and Validation of the Machine Learning Model

After going through the hyperparameter grid and the SMOTE grid, the best model was selected according to the MCC metric.

The corresponding hyperparameters, k-nearest neighbors in the SMOTE, and values of MCC on the training dataset are listed in **Table 2**. One might see that the SVM model has the best performance with an average MCC score at 0.666 ± 0.093.With the fixed hyperparameters, the performance of all six ML diagnostic models on the test dataset is shown in **Figure 6**. The trained SVM had the highest accuracy with 89.1%, while the RF showed the lowest but a close accuracy with 85.3% (**Figure 6A**), which demonstrated the robustness of both hub genes and ML methods.

Based on the results, the weakest performance of the sensitivity metric was the KNN algorithm with a ratio of 81.7% (**Figure 6A**). As a contrast, the SVM algorithm again had the highest sensitivity of 93.9%, showing the great ability for predicting tumor samples (**Figure 6A**). For specificity, the NN algorithm had the best performance with a 90.5% specificity to predict normal samples. The RF algorithm had the lowest specificity of 82.4%. The SVM algorithm had the second lowest specificity of 83.8% (**Figure 6A**). These results demonstrated the six models have both advantages and disadvantages.

MCC and F1 scores could serve as more reliable metrics which involve all four terms: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) in the confusion matrix. According to the ratios of the MCC and F1 scores, the SVM should be the best model with 78.4 and 89%, respectively (**Figure 6A**).

ROC curves for all six ML classification diagnostic assistants were built using the predicted probability of belonging to different classes. Except for the AUC of DT having the lowest value of 85%, all the other models had an AUC of 93–95% (**Figure 7**).

The prediction performance of six ML diagnostic assistants was further evaluated on the EV dataset (25 tumor samples and 25 normal ones). The results showed the six models can classify all the normal samples correctly with specificity and precision both equaling to 1 (**Figure 8B**); however, the prediction on tumor samples varies. SVM and NN have the best performances on successively predicting 23 tumor samples (**Figure 6D**). As a result, SVM and NN share the highest MCC and F1 scores of 92.3 and 96%, respectively (**Figure 6B**). The AUC of SVM and NN on EV is 99% (**Figures 7A, D**). Therefore, one can conclude that the SVM model based on the expression profiles of three hub genes may have a potential diagnostic value for gastric cancer.

3.5 Semi-Supervised Diagnostic Model

Semi-supervised ML can learn from a combination of small amounts of labeled samples and large amounts of unlabeled ones, which is especially suitable for the scenario of annotating large amounts of samples with expensive costs or miscellaneous steps. In this study, the label spreading (LS) algorithm was tested on 50% (LS50), 60% (LS60), 70% (LS70), 80% (LS80), and 90% (LS90) randomly unlabeled samples in TD. Each learning model was cross-validated 100 times with random permutation. The results shown in **Figure 8** demonstrate that the LS algorithm can successfully learn and predict the sample type even when small amounts of labeled data are available. The mean MCC and F1 scores are 0.649 ± 0.029 and 0.824 ± 0.015, respectively, with 50% unlabeled samples. As the ratio of unlabeled samples increases, the performance of the LS slightly decreases. However, with 90% unlabeled data, the LS90 model still has mean MCC and F1 scores of 0.635 ± 0.020 and 0.816 ± 0.012, respectively. Furthermore, all LS models achieved a good prediction performance for the EV dataset, for example, the LS90 model has mean MCC and F1 scores of 0.845 ± 0.066 and 0.919 ± 0.037, respectively.

4 DISCUSSION AND CONCLUSION

Gastric cancer is still a major disease threatening human health, so it is particularly important to find a comprehensive and effective set of biomarkers with diagnostic values. This study systematically used a series of bioinformatics methods to select key features, i.e., hub genes, which were further confirmed by both, the GEPIA2 tool and IHC experiments. The transcription factor–hub gene regulatory network confirmed that three genes are closely associated with gastric cancer in the level of transcription factors. Based on these features, ML diagnostic assistants for the diagnosis of gastric cancer were established by both supervised and semi-supervised learning. The performance of the ML models on the EV dataset further approves the potential diagnostic ability.

In this study, five GEO datasets were downloaded for construction, and one independent GEO dataset GSE33335 was used for external validation. Comprehensive data collation can make the construction of diagnostic assistants more objective (Ahluwalia et al., 2021; Dai W. et al., 2021; Ye et al., 2021). GSE66229 was used for the WGCNA analysis. WGCNA is a widely used target therapy analysis tool, which clusters related genes, according to some clinical characteristics of research subjects. There have been many studies on gastric cancer tumor markers in recent years, and most of the WGCNA clusterings are based on differentially expressed genes (DEGs) in the research dataset (Li et al., 2021; Xiang et al., 2021; Zhang et al., 2022). In contrast, this study performed the WGCNA analysis on all gastric cancer–related genes in one dataset and fused them with selected features using a supervised learning method, i.e., Fisher score algorithm on five combined datasets, preserving the diversity of the gastric cancer hub genes. This also reduces the hub gene bias caused by clustering with a certain clinical feature traditionally (Yang et al., 2022). We put WGCNA-significant modules into Cytoscape to construct a gene–gene interaction network. Previous research shows that gene–gene interaction networks can reveal the principle and mechanisms of cancer (Zeng et al., 2013; Rana et al., 2020). In order to enhance the objectivity and authenticity of the hub genes, genes that are highly associated with gastric cancer screened by the gene–gene interaction network were intersected with the selected feature genes.

Three hub genes were crucial to the next machine learning-based bioinformatics approach. Although no studies used them as combined biomarkers for gastric cancer diagnosis, some studies have screened these genes in the identification of gastric cancer biomarkers and explored them to a certain extent in the field of human and animal experiments on gastric cancer (Lozano-Pope

et al., 2017; Peng et al., 2020; Liu et al., 2021). ESRRG belongs to the estrogen-related receptor family. In one aspect, it has been classified that ESRRG inhibits the occurrence of gastric cancer by inhibiting the Wnt pathway by activating DY131 (Kang et al., 2018). In another, ESRRG can directly bind to the TFF1 promoter, which is a recognized tumor suppressor and inhibits *Helicobacter pylori* infection (Kang et al., 2021). *Helicobacter pylori* infection is a common cause of chronic atrophic gastritis, which is a precancerous lesion (Rolig et al., 2012). ATP4A and ATP4B belong to a family of P-type cation-transporting ATPases. These two genes belong to the gastric proton pump and are antigens of gastric parietal cells, which are diagnostic markers for immune gastric lesions including atrophic gastritis. Through *in vivo* and *in vitro* experiments in animals and humans, researchers have found that ATP4A and ATP4B were partially or fully methylated in gastric cancer cells. It was also verified that the reactivation and demethylation of ATP4A and ATP4B can effectively inhibit the progression of gastric cancer (Lin et al., 2017; Cao et al., 2020). Hence, ATP4A and ATP4B are important tumor suppressor genes.

Six supervised diagnostic models and one semi-supervised diagnostic model were developed based on different algorithms including SVM, RF, KNN, DT, NN, and XGB (supervised), and LS (semi-supervised). The performance was evaluated by seven metrics, namely, accuracy, specificity, sensitivity, precision, MCC, F1 score, and AUC. All the models were trained through cross-validation and further examined on the EV dataset GSE33335. The results suggested that SVM and LS can serve as the most appropriate algorithm for prediction. For example, LS90 can learn from only 10% of labeled data and achieve 0.906 ± 0.008 and 0.986 ± 0.007 AUC scores for 90% unlabeled data and the EV dataset. Therefore, this study demonstrates the potential ability of the ML diagnostic model created with three hub gene expression profiles of 780 samples.

In recent years, bioinformatics analyses based on machine learning have been popularly used in individual medicine. For example, multi-classifiers and deep neural networks are being applied in cancer research (Huang et al., 2018; Zhang et al., 2021). Comparing to previous studies, our research may be more robust in model development and evaluation. First, we included five datasets with 780 samples in the model development and internal validation. Second, we also used an independent dataset only for external validation. Huang et al. (2018) applied multi-classifiers to select gastric cancer-related miRNAs in one dataset and validate their performance in another two datasets. Huang *et al.* and our team both explored the application of SVM in the diagnosis of gastric cancer. Their SVM diagnostic model's AUC was 95% in the training dataset, which is slightly higher than our corresponding AUC (93%). However, their model achieved a biased performance on the two valid datasets: one was 97%, while the other was less than 80%. Relatively fewer samples in model development may be responsible for this performance. Moreover, their two validation datasets were also involved in biomarker selection; thus, they might not be totally independent. Compared with the WGCNA and network control analyses used in our study to screen potential cancer-related genes, Zhang et al. (2021) fused gene expression data and DNA methylation data to obtain relatively more biomarkers for training their deep neural networks. On one hand, their study got an extremely high performance in six metrics. The accuracy, precision, recall, F1 score, and AUC value were all around 99%. On the other hand, the absence of an external validation report makes the generalization ability of their study remain unclear.

More several strengths of this study should be emphasized. First of all, data sources in this study come from Asia. Consistency in data sources may strengthen the pertinence of the model. Second, rich data in six datasets are sorted and then integrated into a comprehensive one to build an objective and effective diagnostic model. Third, hub genes selected from three robust methods were used in combination (WGCNA, gene–gene interaction network, and feature gene selection). Fourth, the selected hub genes are multiple-validated by GEPIA2, HPA databases, and transcription factor–hub gene regulatory network, the results of which further confirm the importance of the selected biomarkers. Finally, the diagnostic model is improved with the SMOTE and passes advanced machine learning analysis on an EV dataset and presented more convincing statistical results than previous studies. This study still has some flaws. First, this study deserves to be verified by subsequent independent experiments. Second, although comprehensive bioinformatics analyses were conducted in this study, an in-depth mechanistic study of three hub genes had not been advanced.

Finally, this study systematically established a gastric cancer diagnostic assistant based on multi-database bioinformatics and machine learning analysis. Our results have a moderate effect on auxiliary diagnosis. We expect future research to test the stability of the model.

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

YL and BX designed this research. QC and YW carried out the data analysis. QC wrote the manuscript. BX revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was financed by grants from the National Natural Science Foundation of China No. 11774300.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2022.905523/full#supplementary-material

REFERENCES

Ahluwalia, P., Kolhe, R., and Gahlay, G. K. (2021). The Clinical Relevance of Gene Expression Based Prognostic Signatures in Colorectal Cancer. *Biochimica Biophysica Acta (BBA) - Rev. Cancer* 1875 (2), 188513. doi:10.1016/j.bbcan.2021.188513

Ali, H. E. A., Lung, P.-Y., Sholl, A. B., Gad, S. A., Bustamante, J. J., Ali, H. I., et al. (2018). Dysregulated Gene Expression Predicts Tumor Aggressiveness in African-American Prostate Cancer Patients. *Sci. Rep.* 8 (1), 16335. doi:10.1038/s41598-018-34637-8

Altman, D. G., and Bland, J. M. (1994). Statistics Notes: Diagnostic Tests 1: Sensitivity and Specificity. *BMJ* 308 (6943), 1552. doi:10.1136/bmj.308.6943.1552

Asplund, J., Kauppila, J. H., Mattsson, F., and Lagergren, J. (2018). Survival Trends in Gastric Adenocarcinoma: A Population-Based Study in Sweden. *Ann. Surg. Oncol.* 25 (9), 2693–2702. doi:10.1245/s10434-018-6627-y

Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T., and Albrecht, M. (2007). Computing Topological Parameters of Biological Networks. *Bioinformatics* 24 (2), 282–284. doi:10.1093/bioinformatics/btm554

Boucher, B., and Jenna, S. (2013). Genetic Interaction Networks: Better Understand to Better Predict. *Front. Genet.* 4, 290. doi:10.3389/fgene.2013.00290

Byvatov, E., and Schneider, G. (2003). Support Vector Machine Applications in Bioinformatics. *Appl. Bioinforma.* 2 (2), 67–77.

Cao, D., Zhao, D., Jia, Z., Su, T., Zhang, Y., Wu, Y., et al. (2020). Reactivation of Atp4a Concomitant with Intragenic DNA Demethylation for Cancer Inhibition in a Gastric Cancer Model. *Life Sci.* 242, 117214. doi:10.1016/j.lfs.2019.117214

Carvalho, B. S., and Irizarry, R. A. (2010). A Framework for Oligonucleotide Microarray Preprocessing. *Bioinformatics* 26 (19), 2363–2367. doi:10.1093/bioinformatics/btq431

Chapelle, O., Scholkopf, B., and Zien, Eds., A. (2009). Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews]. *IEEE Trans. Neural Netw.* 20 (3), 542. doi:10.1109/tnn.2009.2015974

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* 16, 321–357. doi:10.1613/jair.953

Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, August 13–17, 2016 (San Francisco, CA: Association for Computing Machinery).

Chen, X., and Ishwaran, H. (2012). Random Forests for Genomic Data Analysis. *Genomics* 99 (6), 323–329. doi:10.1016/j.ygeno.2012.04.003

Chen, X., Wang, M., and Zhang, H. (2011). The Use of Classification Trees for Bioinformatics. *WIREs Data Min. Knowl. Discov.* 1 (1), 55–63. doi:10.1002/widm.14

Chicco, D., and Jurman, G. (2020). The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics* 21 (1), 6. doi:10.1186/s12864-019-6413-7

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: an R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res.* 44 (8), e71. doi:10.1093/nar/gkv1507

Dai, W., Feng, J., Hu, X., Chen, Y., Gu, Q., Gong, W., et al. (2021a). SLC7A7 Is a Prognostic Biomarker Correlated with Immune Infiltrates in Non-small Cell Lung Cancer. *Cancer Cell Int.* 21 (1), 106. doi:10.1186/s12935-021-01781-7

Dai, Y., Yang, G., Yang, L., Jiang, L., Zheng, G., Pan, S., et al. (2021b). Expression of FOXA1 Gene Regulates the Proliferation and Invasion of Human Gastric Cancer Cells. *Cell Mol. Biol. (Noisy-le-grand)* 67 (2), 161–165. doi:10.14715/cmb/2021.67.2.25

Farhadian, M., Rafat, S. A., Panahi, B., and Mayack, C. (2021). Weighted Gene Co-Expression Network Analysis Identifies Modules and Functionally Enriched Pathways in the Lactation Process. *Sci. Rep.* 11 (1), 2367. doi:10.1038/s41598-021-81888-z

Fuxman Bass, J. I., Diallo, A., Nelson, J., Soto, J. M., Myers, C. L., and Walhout, A. J. M. (2013). Using Networks to Measure Similarity between Genes: Association Index Selection. *Nat. Methods* 10 (12), 1169–1176. doi:10.1038/nmeth.2728

Gao, Y., Zhang, S., Zhang, Y., and Qian, J. (2020). Identification of MicroRNA-Target Gene-Transcription Factor Regulatory Networks in Colorectal Adenoma Using Microarray Expression Data. *Front. Genet.* 11, 463. doi:10.3389/fgene.2020.00463

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy--Analysis of Affymetrix GeneChip Data at the Probe Level. *Bioinformatics* 20 (3), 307–315. doi:10.1093/bioinformatics/btg405

Gu, Q., Li, Z., and Han, J. (2011). "Generalized Fisher Score for Feature Selection," in Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (Barcelona, Spain: AUAI Press).

Heidaryan, E. (2018). A Note on Model Selection Based on the Percentage of Accuracy-Precision. *J. Energy Resour. Technol.* 141 (4), 045501. doi:10.1115/1.4041844

Hu, H.-M., Tsai, H.-J., Ku, H.-Y., Lo, S.-S., Shan, Y.-S., Chang, H.-C., et al. (2021). Survival Outcomes of Management in Metastatic Gastric Adenocarcinoma Patients. *Sci. Rep.* 11 (1), 23142. doi:10.1038/s41598-021-02391-z

Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., and Sun, Q. (2018). Deep Learning for Image-Based Cancer Detection and Diagnosis − A Survey. *Pattern Recognit.* 83, 134–149. doi:10.1016/j.patcog.2018.05.014

Huang, Y., Zhu, J., Li, W., Zhang, Z., Xiong, P., Wang, H., et al. (2018). Serum microRNA Panel Excavated by Machine Learning as a Potential Biomarker for the Detection of Gastric Cancer. *Oncol. Rep.* 39 (3), 1338–1346. doi:10.3892/or.2017.6163

Janky, R. S., Verfaillie, A., Imrichová, H., Van de Sande, B., Standaert, L., Christiaens, V., et al. (2014). iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput. Biol.* 10 (7), e1003731. doi:10.1371/journal.pcbi.1003731

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* 8 (1), 118–127. doi:10.1093/biostatistics/kxj037

Kang, M.-H., Choi, H., Oshima, M., Cheong, J.-H., Kim, S., Lee, J. H., et al. (2018). Estrogen-Related Receptor Gamma Functions as a Tumor Suppressor in Gastric Cancer. *Nat. Commun.* 9 (1), 1920. doi:10.1038/s41467-018-04244-2

Kang, M.-H., Eyun, S.-I., and Park, Y.-Y. (2021). Estrogen-Related Receptor-Gamma Influences *Helicobacter P* Infection by Regulating TFF1 in Gastric Cancer. *Biochem. Biophys. Res. Commun.* 563, 15–22. doi:10.1016/j.bbrc.2021.05.076

Katai, H., Ishikawa, T., Ishikawa, T., Akazawa, K., Isobe, Y., Miyashiro, I., et al. (2018). Five-Year Survival Analysis of Surgically Resected Gastric Cancer Cases in Japan: A Retrospective Analysis of More Than 100,000 Patients from the Nationwide Registry of the Japanese Gastric Cancer Association (2001-2007). *Gastric Cancer* 21 (1), 144–154. doi:10.1007/s10120-017-0716-7

Lancashire, L. J., Lemetre, C., and Ball, G. R. (2009). An Introduction to Artificial Neural Networks in Bioinformatics-Aapplication to Complex Microarray and Mass Spectrometry Datasets in Cancer Studies. *Briefings Bioinforma.* 10 (3), 315–329. doi:10.1093/bib/bbp012

Langfelder, P., and Horvath, S. (2008). WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Li, C., Hou, X., Yuan, S., Zhang, Y., Yuan, W., Liu, X., et al. (2021). High Expression of TREM2 Promotes EMT via the PI3K/AKT Pathway in Gastric Cancer: Bioinformatics Analysis and Experimental Verification. *J. Cancer* 12 (11), 3277–3290. doi:10.7150/jca.55077

Lin, M., Pan, J., Chen, Q., Xu, Z., Lin, X., and Shi, C. (2018). Overexpression of FOXA1 Inhibits Cell Proliferation and EMT of Human Gastric Cancer AGS Cells. *Gene* 642, 145–151. doi:10.1016/j.gene.2017.11.023

Lin, S., Lin, B., Wang, X., Pan, Y., Xu, Q., He, J.-S., et al. (2017). Silencing of ATP4B of ATPase H+/K+ Transporting Beta Subunit by Intragenic Epigenetic Alteration in Human Gastric Cancer Cells. *Oncol. Res.* 25 (3), 317–329. doi:10.3727/096504016X14734735156265

Liu, J., Feng, W., Liu, M., Rao, H., Li, X., Teng, Y., et al. (2021). Stomach-Specific C-Myc Overexpression Drives Gastric Adenoma in Mice via AKT/mTOR Signaling. *Bosn J Basic Med Sci* 21 (4), 434–446. doi:10.17305/bjbms.2020.4978

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* 45 (6), 580–585. doi:10.1038/ng.2653

Lozano-Pope, I., Sharma, A., Matthias, M., Doran, K. S., and Obonyo, M. (2017). Effect of Myeloid Differentiation Primary Response Gene 88 on Expression Profiles of Genes during the Development and Progression of Helicobacter-

Induced Gastric Cancer. *BMC cancer* 17 (1), 133. doi:10.1186/s12885-017-3114-y

Peng, Z., Guan, Q., Luo, J., Deng, W., Liu, J., Yan, R., et al. (2020). Sophoridine Exerts Tumor-Suppressive Activities via Promoting ESRRG-Mediated β-Catenin Degradation in Gastric Cancer. *BMC Cancer* 20 (1), 582. doi:10.1186/s12885-020-07067-x

Rana, H. K., Akhtar, M. R., Islam, M. B., Ahmed, M. B., Lió, P., Huq, F., et al. (2020). Machine Learning and Bioinformatics Models to Identify Pathways that Mediate Influences of Welding Fumes on Cancer Progression. *Sci. Rep.* 10 (1), 2795. doi:10.1038/s41598-020-57916-9

Rolig, A. S., Shanks, J., Carter, J. E., and Ottemann, K. M. (2012). *Helicobacter P* Requires TlpD-Driven Chemotaxis to Proliferate in the Antrum. *Infect. Immun.* 80 (10), 3713–3720. doi:10.1128/IAI.00407-12

Royston, P., Parmar, M. K. B., and Sylvester, R. (2004). Construction and Validation of a Prognostic Model across Several Studies, with an Application in Superficial Bladder Cancer. *Stat. Med.* 23 (6), 907–926. doi:10.1002/sim.1691

Saragoni, L., Morgagni, P., Gardini, A., Marfisi, C., Vittimberga, G., Garcea, D., et al. (2013). Early Gastric Cancer: Diagnosis, Staging, and Clinical Impact. Evaluation of 530 Patients. New Elements for an Updated Definition and Classification. *Gastric Cancer* 16 (4), 549–554. doi:10.1007/s10120-013-0233-2

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-Omics Data Integration, Interpretation, and its Application. *Bioinform Biol. Insights* 14, 1177932219899051. doi:10.1177/1177932219899051

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tang, Z., Kang, B., Li, C., Chen, T., and Zhang, Z. (2019). GEPIA2: An Enhanced Web Server for Large-Scale Expression Profiling and Interactive Analysis. *Nucleic Acids Res.* 47 (W1), W556–W560. doi:10.1093/nar/gkz430

Thul, P. J., and Lindskog, C. (2018). The Human Protein Atlas: A Spatial Map of the Human Proteome. *Protein Sci.* 27 (1), 233–244. doi:10.1002/pro.3307

Van Cutsem, E., Sagaert, X., Topal, B., Haustermans, K., and Prenen, H. (2016). Gastric Cancer. *Lancet* 388 (10060), 2654–2664. doi:10.1016/S0140-6736(16)30354-3

Xiang, P., Liu, D., Guan, D., Du, Z., Hao, Y., Yan, W., et al. (2021). Identification of Key Genes in Benign Prostatic Hyperplasia Using Bioinformatics Analysis. *World J. Urol.* 39 (9), 3509–3516. doi:10.1007/s00345-021-03625-5

Yang, A., Zhang, W., Wang, J., Yang, K., Han, Y., and Zhang, L. (2020). Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA. *Front. Bioeng. Biotechnol.* 8, 1032. doi:10.3389/fbioe.2020.01032

Yang, H., Tian, W., and Zhou, B. (2022). Sarcopenia and a 5-mRNA Risk Module as a Combined Factor to Predict Prognosis for Patients with Stomach Adenocarcinoma. *Genomics* 114 (1), 361–377. doi:10.1016/j.ygeno.2021.12.011

Ye, W., Luo, C., Liu, F., Liu, Z., and Chen, F. (2021). CD96 Correlates with Immune Infiltration and Impacts Patient Prognosis: A Pan-Cancer Analysis. *Front. Oncol.* 11, 634617. doi:10.3389/fonc.2021.634617

Zeng, T., Sun, S.-Y., Wang, Y., Zhu, H., and Chen, L. (2013). Network Biomarkers Reveal Dysfunctional Gene Regulations during Disease Progression. *FEBS J.* 280 (22), 5682–5695. doi:10.1111/febs.12536

Zhang, G., Liu, X., Sun, Z., Feng, X., Wang, H., Hao, J., et al. (2022). A2M Is a Potential Core Gene in Intrahepatic Cholangiocarcinoma. *BMC Cancer* 22 (1), 5. doi:10.1186/s12885-021-09070-2

Zhang, G., Xue, Z., Yan, C., Wang, J., and Luo, H. (2021). A Novel Biomarker Identification Approach for Gastric Cancer Using Gene Expression and DNA Methylation Dataset. *Front. Genet.* 12, 644378. doi:10.3389/fgene.2021.644378

Zhang, Z. (2016). Introduction to Machine Learning: K-Nearest Neighbors. *Ann. Transl. Med.* 4 (11), 218. doi:10.21037/atm.2016.03.37

Zhou, D., Bousquet, O., Lal, T., Weston, J., and Schölkopf, B. (2003). Learning with Local and Global Consistency. *Adv. Neural Inf. Process. Syst.* 16, 321–328.

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, Wang, Liu and Xi. 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 Physiology

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fphys.2022.918929

Multi-Model Domain Adaptation for Diabetic Retinopathy Classification

Guanghua Zhang[1,2†], *Bin Sun*[3†], *Zhaoxia Zhang*[3†], *Jing Pan*[4], *Weihua Yang*[5*] and *Yunfang Liu*[6*]

[1]Department of Intelligence and Automation, Taiyuan University, Taiyuan, China, [2]Graphics and Imaging Laboratory, University of Girona, Girona, Spain, [3]Shanxi Eye Hospital, Taiyuan, China, [4]Department of Materials and Chemical Engineering, Taiyuan University, Taiyuan, China, [5]Affiliated Eye Hospital, Nanjing Medical University, Nanjing, China, [6]The First Affiliated Hospital of Huzhou University, Huzhou, China

Diabetic retinopathy (DR) is one of the most threatening complications in diabetic patients, leading to permanent blindness without timely treatment. However, DR screening is not only a time-consuming task that requires experienced ophthalmologists but also easy to produce misdiagnosis. In recent years, deep learning techniques based on convolutional neural networks have attracted increasing research attention in medical image analysis, especially for DR diagnosis. However, dataset labeling is expensive work and it is necessary for existing deep-learning-based DR detection models. For this study, a novel domain adaptation method (multi-model domain adaptation) is developed for unsupervised DR classification in unlabeled retinal images. At the same time, it only exploits discriminative information from multiple source models without access to any data. In detail, we integrate a weight mechanism into the multi-model-based domain adaptation by measuring the importance of each source domain in a novel way, and a weighted pseudo-labeling strategy is attached to the source feature extractors for training the target DR classification model. Extensive experiments are performed on four source datasets (DDR, IDRiD, Messidor, and Messidor-2) to a target domain APTOS 2019, showing that MMDA produces competitive performance for present state-of-the-art methods for DR classification. As a novel DR detection approach, this article presents a new domain adaptation solution for medical image analysis when the source data is unavailable.

Keywords: diabetic retinopathy classification, multi-model, domain adaptation, convolutional neural network, deep learning

OPEN ACCESS

Edited by:
Wei Wang,
Huzhou Maternity and Child Health Care Hospital, China

Reviewed by:
Xuefei Song,
Shanghai Ninth People's Hospital, China
Somayah Abdullah Albaradei,
King Abdullah University of Science and Technology, Saudi Arabia

***Correspondence:**
Yunfang Liu
panlele0701@139.com
Weihua Yang
benben0606@139.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

Received: *12 April 2022*
Accepted: *12 May 2022*
Published: *01 July 2022*

Citation:
Zhang G, Sun B, Zhang Z, Pan J, Yang W and Liu Y (2022) Multi-Model Domain Adaptation for Diabetic Retinopathy Classification. Front. Physiol. 13:918929. doi: 10.3389/fphys.2022.918929

1 INTRODUCTION

Diabetic retinopathy (DR) is a complication of diabetic patients and a significant cause of blindness globally among the working population (Antonetti et al., 2021). There are 451 million suffering from DR in the world, and this is projected to increase to 639 million in 2045 (Cho et al., 2018). In diabetics, blood is provided to all retina layers through micro blood vessels that are sensitive to unrestricted blood sugar levels. DR may cause no symptoms or only mild vision problems at first, but it can cause blindness eventually. When substantial glucose or fructose is collected in the blood, blood vessels begin to collapse due to insufficient oxygen supply to the cells. Occlusion in these blood vessels can cause serious eye damage. As a result, metabolic rate decreases, and abnormal blood vessels accumulate in DR (Dai et al., 2021). Microaneurysms (MAs) are the early signs of DR, which

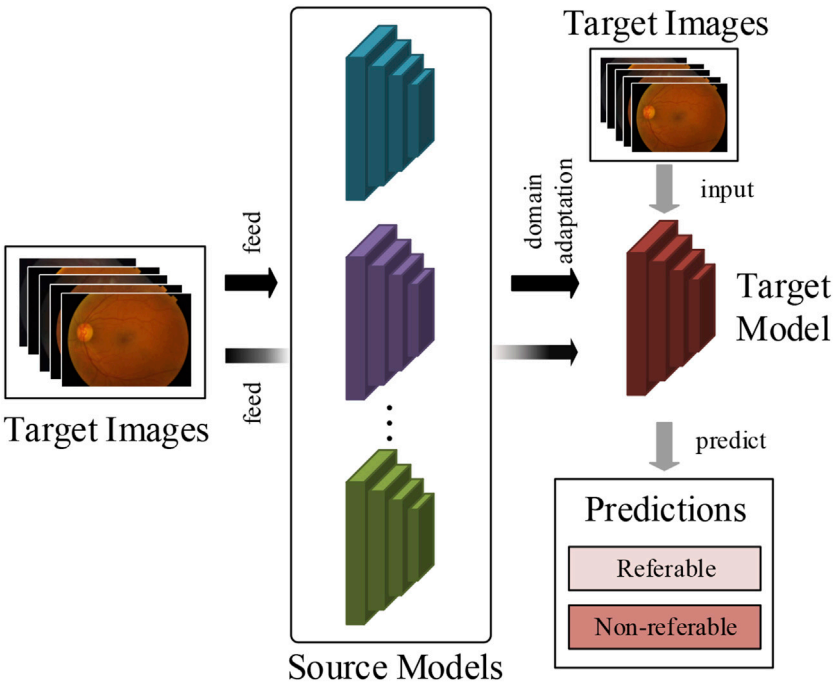


FIGURE 1 | The work flow of our method. We train the target prediction model to simply use pre-trained multiple source models and unlabeled target retinal images.

cause changes in the size (swelling) of the blood vessels. Moreover, hemorrhages (HMs), exudates (EXs), and abnormal blood vessel growth are the symptoms of DR. The International Clinical Diabetic Retinopathy (ICDR) scale is one of the most commonly used clinical scales and is composed of five levels of DR: normal, mild, moderate, severe and proliferative (Bodapati et al., 2021). Generally, diabetic retinopathy is divided into referable diabetic retinopathy (RDR) and non-referable diabetic retinopathy (NRDR).

Blindness can be completely avoided by early diagnosis. Annual regular clinical examination for diabetics is strongly recommended, especially for middle-aged and older adults (Mohamed et al., 2007; Ferris, 1993). Nevertheless, researchers find that a considerable number of people with diabetes failed to have annual eye examinations due to very mild symptoms, long examination time, and a shortage of ophthalmologists (Owsley et al., 2006; MacLennan et al., 2014; Chou et al., 2014). Therefore, it is necessary to adopt automatic DR diagnosis methods to lighten the workload on eye specialists and shorten the detection time, making patients understand the condition and get treatment in time.

Artificial intelligence (AI) is a popular technique for computer-aided automatic DR diagnosis to overcome these obstacles and deep learning has achieved progress in biomedical image analysis (Meng et al., 2021b; Preston et al., 2021; Meng et al., 2021a). Yoo and Park (2013) utilized ridge, elastic net, and LASSO to perform validation on 1052 DR patients. Roychowdhury et al. (2013) proposed a novel two-step approach for DR detection, where non-lesions or normal images are rejected in the first step, and bright and red lesions are classified as hard exudates and hemorrhages, respectively in the second step. In addition to the machine learning methods, the deep learning method becomes very popular in DR screening in recent years. For instance, Vo and Verma (2016) used a deep neural network improved upon GoogLeNet and VGGNet for DR recognition, aiming to learn fine-grained features of retinal images. Moreover, He et al. (2020) combined two attention blocks with a backbone network to solve the imbalanced DR data distribution problem and capture more detailed lesion information, respectively. Ai et al. (2021) proposed an algorithm adopting deep ensemble learning and attention mechanism to detect DR. However, both traditional machine learning methods (Yoo and Park, 2013; Roychowdhury et al., 2013) and supervised deep learning methods (Vo and Verma, 2016; He et al., 2020; Mohamed et al., 2021; Ai et al., 2021) require a large amount of labeled retinal images to train their models, which fail to new data from other domains. As an effective solution, domain adaptation always requires source data, which is usually difficult to access in practical applications because of the strict privacy rules in medical image management agencies.

To tackle this critical problem in supervised deep learning methods, this article attempts to develop a multi-model domain adaptation (MMDA) to conduct transfer learning for DR classification without access to source data. As shown in **Figure 1**, the proposed method can sufficiently utilize the knowledge of source models and unlabeled target images to improve the DR detection performance.

In the MMDA framework, the target model is initially parameterized, and the trained source models are provided. We propose a model weight determination module to estimate

the importance of each source model by measuring the average distance between two retinal feature groups extracted from the source models and target model. This module is optimized by a weight determination loss to output realistic model weights in target feature learning. By using the weights of source models, the pseudo label of the target images is obtained in a feature-level clustering-based way. Finally, we optimize the target model by cross-entropy loss and information maximization loss to guarantee the performance of diabetic retinopathy detection.

To evaluate the performance of MMDA, we conduct extensive experiments on five publicly available retinal image datasets: DDR, IDRiD, Messidor, Messidor-2, and APTOS 2019, obtaining excellent performance without access to source data. The results demonstrate that our proposed method can effectively complete the DR diagnosis task with only unlabeled target data.

2 MATERIALS AND METHODS

2.1 Data Acquisition

In order to validate our method for diabetic retinopathy diagnosis, we trained four source models from publicly available datasets (DDR, IDRiD, Messidor, and Messidor-2) and employed APTOS 2019 as the target domain.

DDR dataset (Diabetic retinopathy detection, 2015) involves 12,522 fundus images from a 45° field of view. In detail, it has 6,266 normal fundus images and 6,256 abnormal samples. Moreover, the class distribution of the dataset is imbalanced in that the normal images are more than the abnormal data.

The IDRiD dataset (Porwal et al, 2018) contains 516 fundus images which were captured by an ophthalmologist from an Indian eye clinic. It provides adequate quality and clinically relevant fundus images with ground truths.

The Messidor dataset (Decencière et al., 2014) is a publicly available diabetic retinopathy dataset provided by the Messidor program partners, which consists of 1,200 retinal images, and for each image, two grades, retinopathy grade, and risk of macular edema, are provided.

The Messidor-2 dataset (Decencière et al., 2014) has been globally used by researchers for DR detection algorithm analysis, which is an extension of Messidor. It contains 1,748 retinal images of 874 examinations. Although there are no official annotations for this dataset, the third-party grades for 1,744 out of the 1,748 images adjudicated by a panel of three retina specialists are available for researchers (Messidor-2 dr grades, 2018).

The APTOS 2019 dataset (Khalifa et al., 2019) is the most recent publicly available Kaggle dataset from the APTOS Blindness Detection competition on Kaggle for DR detection. It contains 3,662 labeled fundus photography images.

The above datasets are graded into five stages from 0 to 4 for no DR, mild DR, moderate DR, severe DR, and proliferative DR, respectively, according to the ICDR severity scale. The label distribution of the datasets and the division of the referable and non-referable DR are shown in **Table 1**. Moreover, the APTOS 2019 dataset is regarded as the target domain, and the other four datasets are used as source datasets to train source models.

TABLE 1 | Label distributions of DDR, IDRiD, Messidor, Messidor-2, and APTOS 2019 datasets.

Dataset	Type	Non-referable			Referable	
		No	Mild	Moderate	Severe	Proliferative
DDR	Source	6,266	630	4,477	236	913
IDRiD	Source	168	25	168	93	62
Messidor	Source	546	153	247	254	—
Messidor-2	Source	1,017	270	347	75	35
APTOS 2019	Target	1,805	370	999	193	295

2.2 Data Preprocessing

When collecting retinal images, the differences in lighting conditions and camera types may cause a large data inconsistency (Graham, 2015). Data preprocessing mitigates noise and enhances image details, reducing inconsistency and playing a significant role in improving performance.

In order to eliminate these negative effects and make data consistent, we perform data preprocessing in the following two steps (**Figure 2**):

Step 1: Resize and crop.

Considering various resolutions of retinal images in different datasets, we resize all images to 1,024 pixels if their width or height is bigger than that size. Then, we crop as much of the black space by identifying the center and radius of the circle in the retinal images.

Step 2: Image enhancement.

In DR detection, the observation of hard exudates, hemorrhages, and cotton wools is significant for eye specialists to diagnose. However, the variations of brightness and resolution not only make ophthalmologists produce misdiagnoses but also make it difficult for a model to compose robust features. To address this problem, we perform image enhancement after resizing and cropping by the following formula:

$$I_o(x, y; \sigma) = \lambda G(x, y; \sigma) * I(x, y) + \omega I(x, y) + \delta \quad (1)$$

where $I(x, y)$ denotes the input retinal image, $G(x, y; \sigma)$ is a Gaussian filter with standard deviation σ, "$*$" represents the convolution operator. λ, ω and δ are manually adjusted variables. In our study, λ, ω and δ are set to 4, -4 and 128, respectively. By improving image contrast with **Eq. 1**, the lesion area is easier to distinguish.

2.3 Multi-Model Domain Adaptation Architecture

This subsection elaborates on our proposed MMDA method, which aims to address the central problem that the labeled image data cannot always be obtained in automatic DR detection.

2.3.1 Overview

Domain adaptation is one of the branches of transfer learning in computer science. For a vanilla multi-source unsupervised domain adaptation task, we have n source domains with fundus images, and N_s^i labeled samples $\{x_i^j, y_i^j\}_{j=1}^{N_s^i}$ from the ith source domain are given, where $x_i^j \in \mathcal{X}_i$, $y_i^j \in \mathcal{Y}_i$, and also N_t unlabeled retinal images $\{x_t^j\}_{j=1}^{N_t}$

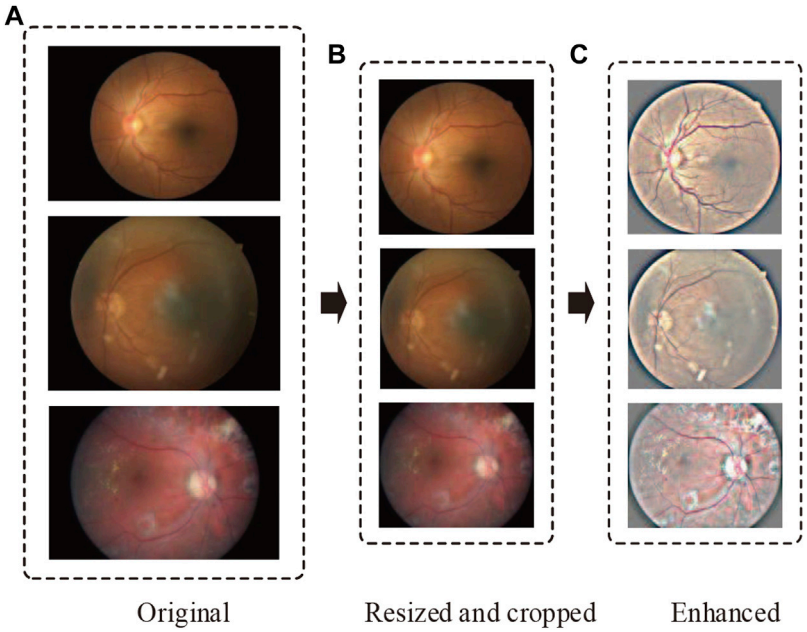


FIGURE 2 | Representative retinal images adopting our preprocessing techniques. From top to bottom, the representative images are sampled from no DR, moderate DR, and proliferative DR, respectively. The parts **(A–C)** denote the original, resized and cropped, and enhanced retinal images.

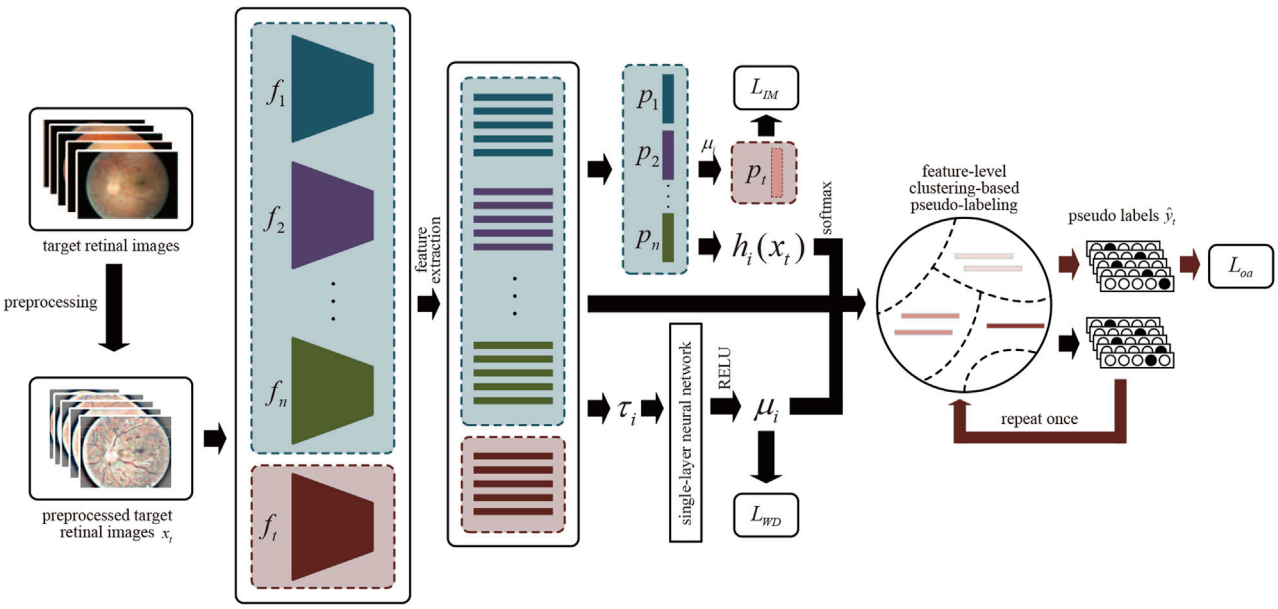


FIGURE 3 | The overview of MMDA architecture. After preprocessing, we obtain the features of target retinal images by source models $f_i|_{i=1}^n$ and target model f_t, and calculate the weight of each model μ_i using the single-layer neural network. The output of the target classifier is defined by the source classifiers with fixed parameters. Pseudo labels $\hat{y}_t$ for each retinal image x_t are obtained after the process of feature-level clustering-based pseudo-labeling.

from the target domain $\mathcal{D}_t$ where $x_t^j \in \mathcal{X}_t$. Domain adaptation aims to obtain a target model to predict the labels $\{y_t^j\}_{j=1}^{N_t}$, where $y_t^j \in \mathcal{Y}_t$. Here, the goal of MMDA is to learn a target prediction model for function h_t: $\mathcal{X}_t \to \mathcal{Y}_t$ and infer $\{y_t^i\}_{i=1}^{N_t}$, with only $\{x_t^i\}_{i=1}^{N_t}$ and the source prediction models for function: h_i: $\mathcal{X}_i \to \mathcal{Y}_i$ available. Note that, only trained source models can be utilized, without access to

their data. **Figure 3** illustrates the overview of our MMDA model for referable DR detection.

Suppose that we have multiple trained source models for DR classification and an ImageNet pre-trained target model. Each model contains two modules: the feature encoding module f_i: $\mathcal{X}_i \to \mathbb{R}^d$ and the classifier module p_i: $\mathbb{R}^d \to \mathbb{R}^K$, i.e., $h_i(x) = p_i\,(f_i(x))$. Here, d is the dimension of the feature, and K represents the number of categories. We extract the features of retinal images using the source and target feature encoding modules firstly. A single-layer neural network is integrated to determine the weights of each source model using euclidean distance and weight determination loss. By employing the weights and source classifiers, target prediction logits are obtained. Then the pseudo labels $\hat{y}_t$ of the target images are generated in a feature-level clustering-based way. Finally, the whole network is optimized using cross-entropy loss and information maximization loss to make the target feature encoding module has excellent DR diagnosis capability.

2.3.2 Source Model Generation

We consider producing several source backbone pre-trained source models, i.e., $h_i = p_i \circ f_i$ ($i = 1, 2, \cdots n$), by optimizing them using the following cross-entropy loss:

$$L^i_{src}\left(f_i; \mathcal{X}_i, \mathcal{Y}_i\right) = -\mathbb{E}_{\left(x_i, y_i\right)\in\mathcal{X}_i\times\mathcal{Y}_i} \sum_{k=1}^{K} q_k \log\, \sigma_k\left(h_i\left(x_i\right)\right) \quad (2)$$

where $\sigma_k(a) = \frac{\exp(a_k)}{\sum_i \exp(a_i)}$ denotes the kth element in the softmax output of a K-dimensional vector a, and q is the one-hot encoding of y_i where q_k is set to "1" if y_i is the kth class and the rest is set to "0". In order to learn more discriminative feature representations and further enhance the following target data alignment, we adopt the label smoothing technique because it prevents the model from becoming over-confident thus improving generalization and performance (Müller et al., 2019). With label smoothing, the objective loss function is modified as below:

$$L^i_{src,ls}\left(f_i; \mathcal{X}_i, \mathcal{Y}_i\right) = -\mathbb{E}_{\left(x_i, y_i\right)\in\mathcal{X}_i\times\mathcal{Y}_i} \sum_{k=1}^{K} q_{k,ls} \log\, \sigma_k\left(h_i\left(x_i\right)\right) \quad (3)$$

where $q_{k,ls} = (1 - \alpha)q_k + \alpha/K$ represents the smoothed label and α is the smoothing factor which is set to 0.1 experientially.

2.3.3 Information Maximization Loss for Target Model

Due to the source classifier modules encoding the distribution information of unseen source data, our framework is proposed to learn the domain-specific feature encoding module while the source classifier modules are fixed. Specifically, MMDA employs the weighted source classifier modules during the target model learning process:

$$p_t(\cdot) = \sum_{i=1}^{n} \mu_i p_i(\cdot) \quad (4)$$

where μ_i is the weight for the ith source model, which will be explained in detail in the following subsection.

In essence, our goal is to obtain an optimal target feature extractor f_t: $\mathcal{X}_t \to \mathbb{R}^d$ for target retinal images, in order that the extracted target features can match source distributions well. However, it is noteworthy that the source images are not accessible in our study. As a result, there's no way to perform feature-level alignment since it is unfeasible to estimate the source distribution in the absence of source data. We look at the problem from a different angle that the expected output logits of the target model should seem like one-hot encoding but differ from each other if the domain gaps are mitigated. To this end, we employ the information maximization (IM) loss (Hu et al., 2017), which enhance the certainty and diversity of target outputs. Specifically, we optimize f_t by IM loss L_{IM} that consists of two objective functions L_{ce} and L_{div}:

$$\begin{aligned} L_{ce}\left(f_t; \mathcal{X}_t\right) &= -\mathbb{E}_{x_t\in\mathcal{X}_t} \sum_{k=1}^{K} \sigma_k\left(h_t\left(x_t\right)\right) \log\, \sigma_k\left(h_t\left(x_t\right)\right) \\ L_{div}\left(f_t; \mathcal{X}_t\right) &= \sum_{k=1}^{K} P_k \log\, P_k = D_{KL}\left(P, \frac{1}{K}\mathbf{1}_K\right) - \log\, K \\ L_{IM} &= Lce + \beta L_{div} \end{aligned} \quad (5)$$

where $h_t(x) = p_t\,(f_t(x))$ is the K-dimensional output logits of each retinal images, $\mathbf{1}_K$ is an all-ones vector with K elements, and $P = \mathbb{E}_{x_t\in\mathcal{X}_t}[\sigma(f_t(x_t))]$ represents the average output probabilities of the whole target domain, β is the balance factor. Information maximization would work better than conditional entropy minimization (Grandvalet and Bengio, 2005) commonly used in traditional domain adaptation works, since it can circumvent the trivial solution where all unlabeled fundus images have the same one-hot encoding via the fair diversity-promoting objective function L_{div}.

2.3.4 Model Weight Determination

In the MMDA framework, a robust target feature encoder is learned by bridging the domain gap between each source domain and the target domain. However, the feature discrepancies between each source domain and target domain are different. To measure the feature discrepancies, we propose a Model Weight Mechanism (MWM). Precisely, we first calculate the average Euclidean Distance between the ith source domain and the target domain:

$$\tau_i = \frac{1}{N_t} \sum_{j=1}^{N_t} \left| f_i\left(x_j\right) - f_t\left(x_j\right) \right| \quad (6)$$

The closer the τ_i is, the more important the source model, i.e., the greater the weight is. To this end, we integrate a single-layer neural network, which is parameterized by a weight vector $w = (w_1, w_2, \cdots w_d) \in \mathbb{R}^{d\times 1}$. Formally, we learn model weight μ_i as

$$\mu_i = \frac{\exp\left(\mathrm{ReLU}\left(w^T\tau_i\right)\right)}{\sum_{j=1}^{N} \exp\left(\mathrm{ReLU}\left(w^T\tau_j\right)\right.} \quad (7)$$

where $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$ is an activation function, which guarantees the nonnegativity of μ_i. The role of the above softmax

operation is to guarantee the model weight satisfy the following property:

$$\sum_{i=1}^{n} \mu_i = 1, \mu_i \geq 0 \tag{8}$$

We optimize the weight vector of the single-layer neural network w by minimizing the following loss function:

$$L_{WD}(w; \mathcal{X}_t) = \frac{1}{N_t} \sum_{i=1}^{n} \sum_{j=1}^{N_t} \mu_i \left\| f_i(x_j) - f_t(x_j) \right\|_2^2 \tag{9}$$

That is, a larger distance $\|f_i(x_j) - f_t(x_j)\|_2$ between features $f_i(x)$ and $f_t(x)$ enforces a smaller value μ_i.

2.3.5 Feature-Level Clustering-Based Pseudo-Labeling

The role of IM loss in **Eq. 5** is to enforce the similarity of one-hot encoding output. Therefore, an accurate prediction network is crucial to reduce this impact. For this purpose, a pseudo-labeling strategy at the feature level is applied for better supervision during the adaptation process.

First, the weighted features centroid of target retinal images for each class is obtained, similar to weighted k-means clustering.

$$m_k^{(0)} = \frac{\sum_{i=1}^{n} \sum_{x_t \in \mathcal{X}_t} \mu_i \sigma_k(h_i(x_t)) f_i(x_t)}{\sum_{x_t \in \mathcal{X}_t} \sigma_k(h_i(x_t))} \tag{10}$$

These centroids can robustly and more reliably characterize the distribution of different categories within the target domain. Then, the pseudo labels can be attained via the nearest centroid classifier:

$$\hat{y}_t = \arg\min_k D_c\left(f_t(x_t), m_k^{(0)}\right) \tag{11}$$

where $D_c(a, b) = 1 - \frac{a \cdot b}{\|a\| \|b\|}$ represents the cosine distance between vector a and b.

Based on generated $\hat{y}_t$ previously , new centroids $m_k^{(1)}$ and pseudo labels are computed:

$$\begin{aligned} m_k^{(1)} &= \frac{\sum_{i=1}^{n} \sum_{x_t \in \mathcal{X}_t} 1_{[\hat{y}_t = k]} \mu_i f_i(x_t)}{\sum_{x_t \in \mathcal{X}_t} 1_{[\hat{y}_t = k]}} \\ \hat{y}_t &= \arg\min_k D_c\left(f_t(x_t), m_k^{(1)}\right) \end{aligned} \tag{12}$$

We refer to $\hat{y}_t$ in **Eq. 12** as the final pseudo labels.

To sum up, given n source models $h_i = p_i \circ f_i$ $(i = 1, 2, \cdots n)$ and the final pseudo labels $\hat{y}_t$ generated from **Eq. 12**, MMDA fixes the parameters of sources classifiers, $p_t(\cdot) = \sum_{i=1}^{n} \mu_i p_i(\cdot)$ and optimizes the feature extractor f_t with the overall loss as:

$$L_{oa}(f_t) = L_{IM} - \gamma \mathbb{E}_{(x_t, \hat{y}_t) \in \mathcal{X}_t \times \hat{\mathcal{Y}}_t} \sum_{k=1}^{K} 1_{[k=\hat{y}_t]} \log \sigma_k(p_t(f_t(x_t))) \tag{13}$$

where $\gamma > 0$ is a balancing hyper-parameter. The whole implementation of MMDA model is shown in Algorithm 1.

Algorithm 1. Pseudo-code of MMDA training process

Input: Pre-trained n source models $p_i \circ f_i (i = 1, 2, \cdots n)$, ImageNet pre-trained target model f_t, unlabeled retinal images $\mathcal{X}_t$, β, γ, batch size B, learning rates ζ
Output: target model f_t
Initialize parameters w of the single-layer neural network;
for $epoch = 1$ to N **do**
  Randomly sample x_t of size B from $\mathcal{X}_t$
  for each mini-batch **do**
    Extract features $f_t(x_t)$;
    Calculate L_{IM} and update f_t using Adam optimizer;
    if starting adaptation **then**
      Extract features $f_t(x_t)$ and $f_i(x_t)$ from f_t and $f_i (i = 1, 2, \cdots n)$;
      Calculate μ_i via Eq. (6), Eq. (7), Eq. (9) and update w via L_{WD};
      Calculate L_{IM} via Eq. (4) and Eq. (5);
      Obtain pseudo labels $\hat{y}_t$ via Eq. (10) to Eq.(12);
      Calculate overall loss (Eq. (13)) by obtained pseudo labels;
      Update f_t using Adam optimizer.
    end if
  end for
end for

2.4 Implementation Details

In the experiments we first train the source models by corresponding source retinal datasets, and each model is designed following ResNet-50 (He et al., 2016). As for the target model, it also employs ResNet-50, initialized by pre-trained parameters from ImageNet (Deng et al., 2009). We perform data argumentation by applying random horizontal flips, vertical flips, and random rotation to prevent overfitting. The input size of the MMDA is 224 × 224. We trained 40 epochs for all the source models using the Stochastic Gradient Descent (SGD) optimization algorithm (Kingma and Ba, 2014) with a learning rate decay factor of $1e^{-4}$. The learning rates for DDR, IDRiD, Messidor, and Messidor-2 datasets are 5×10^{-3} equally. For the target training, we adopt a mini-batch SGD with momentum 0.9, weight decay $1e^{-4}$, and learning rate $\zeta = 1e^{-2}$. The balance factor for IM loss β and the overall loss γ are set to 0.3 and 0.3, respectively. In addition, a batch size of 64 is set for the entire experimental process. MMDA is implemented on two NVIDIA RTX 2080Ti GPUs with 2 × 11 GB RAM using the PyTorch framework.

To validate the effectiveness of MWM (**Section 2.3.4**), we adjust μ_i in **Eqs 4**, **10**, **12**. Moreover, hyper-parameters β in L_{IM} and γ are fine-tuned to analyze their influence on DR detection performance. Details are described in **Section 3.3** and **Section 3.4**. Note that we do not integrate model weight mechanism and pseudo-labeling into the target model until training the target model several epochs with IM loss. That means we attach the model weight and pseudo-labeling modules when the target model has a certain diagnosing capability.

3 RESULTS

3.1 Evaluation Metrics

To measure the performance of the MMDA model, we employ accuracy and sensitivity as the measurements. The accuracy can be defined as the percentage of correctly classified images. Sensitivity measures the ability of a test to correctly identify samples with referable DR, which is an effective metric to measure the DR diagnosis.

This metric is calculated as follows. First, we compute the accuracy by $\frac{TP+TN}{TP+FP+TN+FN}$, where TP is the correctly predicted positive samples, TN denotes the correctly predicted negative images, and FP represents the false predicted positive samples and FN means the false predicted negative images. For the sensitivity, it follows the formula,

TABLE 2 | Accuracy and sensitivity of MMDA for diabetic retinopathy diagnosis compared with state-of-the-art supervised learning approaches on the APTOS 2019 dataset.

Method	Accuracy (%)	Sensitivity (%)
Xie et al. (2017)	92.8	86.8
Vives-Boix and Ruiz-Fernández (2021)	94.5	90.0
Narayanan et al. (2020)	98.4	98.9
Farag et al. (2022)	97.0	97.0
MMDA (Ours)	90.6	98.5

TABLE 3 | The DR classification results of MMDA with different backbones on the APTOS 2019 dataset.

Backbone	Method	Accuracy	Sensitivity	δ Accuracy (%)
VGG-16	ACDA	0.873	0.965	↓2.8
	APDA	0.882	0.973	↓1.9
	MMDA	**0.901**	**0.980**	-
ResNet-50	ACDA	0.880	0.972	↓ 2.6
	APDA	0.902	0.960	↓ 0.4
	MMDA	**0.906**	**0.985**	-

The best results are in bold.

TABLE 4 | DR classification results using different β on the APTOS 2019 dataset.

β	0.1	0.2	0.3	0.4	0.5
Accuracy	0.899	0.903	**0.906**	0.902	0.851
Sensitivity	0.984	0.962	**0.985**	0.987	0.986

The best results are in bold.

TABLE 5 | DR classification results using different γ on the APTOS 2019 dataset.

γ	0.1	0.2	0.3	0.4	0.5
Accuracy	0.877	0.896	**0.906**	0.905	0.902
Sensitivity	0.976	0.982	**0.985**	0.966	0.963

The best results are in bold.

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

3.2 Performance Compared With Supervised Learning Methods

We first compare MMDA with the existing supervised learning methods on the APTOS2019 dataset. Specifically, Xie et al. (2017) present a simple, highly modularized network architecture for image classification, which is often employed in DR detection; Vives-Boix and Ruiz-Fernández (2021) conducted automated detection of DR by directly interfering in both learning and memory by reinforcing less common occurrences during the learning process; Narayanan et al. (2020) proposed a hybrid machine learning architecture to detect and grade the level of diabetic retinopathy; Farag et al. (2022) proposed an automatic deep-learning-based model for severity detection by utilizing a single color fundus photograph. From **Table 2**, it is observed that MMDA achieves approving results with 90.6% accuracy and 98.5% sensitivity. Our method performs a relatively excellent accuracy compared with the compared methods, which only remains a distance of 2.2%, and it presents the second-best sensitivity of 98.5%, only weaker than Narayanan et al. (2020). Although the results of MMDA are lower than these supervised learning methods, huge amounts of labeled data are essentially required in their training process. In contrast, we train MMDA simply to utilize unlabeled retinal images and obtain satisfactory performance, showing the superiority of our framework for DR diagnosis.

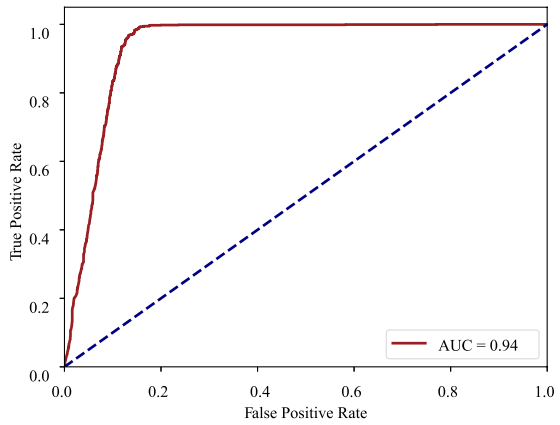


FIGURE 4 | ROC curve of DR diagnosis on the APTOS 2019 dataset.

3.3 Performance Analysis on Model Weight Mechanism

We design a novel model weight mechanism (MWM) to assign a learnable weight to each model. To verify the effect of the MWM, we perform ablation studies to analyze the MWM for the source classifier modules and the pseudo-labeling process using different backbones.

MWM for source classifier modules: We fix the source classifier modules, so we can fully utilize the source distribution information in the modules when the source data is not available. Meanwhile, the discrepancy between each source domain and the target domain cannot be ignored. Specifically, we integrate the weight mechanism into the classifiers by **Eq. 4**. To verify the effect of MWM in utilizing the source distribution information, we conduct a study using Average-weighted Classifier Multi-model Domain Adaptation (ACDA), which is the MMDA model with $\mu_i = \frac{1}{n}\,(i = 1, 2, \cdots n)$ in **Eq. 4**. As shown in **Table 3**, we obtain accuracy and sensitivity of 0.873 and 0.965 for VGG-16, 0.880 and 0.972 for RestNet-50, with accuracy drops of 2.8%, 2.6% for VGG-16 and ResNet-50, respectively, which demonstrates that MWM contributes huge effectiveness on multi-model source distribution learning.

MWM for pseudo-labeling: To evaluate the contribution of MWM in features centroid determination, we carry out an experiment named APDA, which is a modified MMDA with $\mu_i = \frac{1}{n}\,(i = 1, 2, \cdots n)$ in **Eqs 10–12**. With this setting, this model performs at accuracies of 0.882 and 0.902 for VGG-16 and ResNet-50

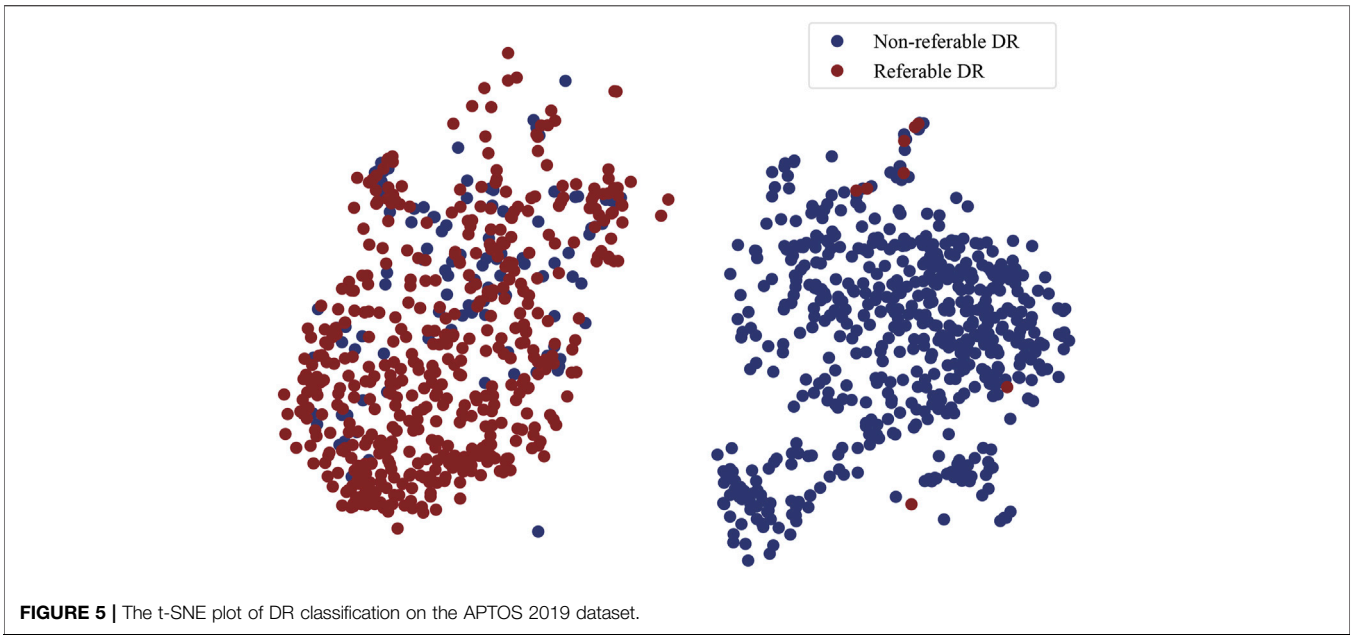


FIGURE 5 | The t-SNE plot of DR classification on the APTOS 2019 dataset.

respectively. All results are lower than the original MMDA model. This is because the average model weight cannot determine accurate centroids of features, which results in incorrect pseudo labels and failing to bridge the domain gap between each source domain and target domain. With MWM, the importance of each source model can be determined, which helps to obtain a more accurate pseudo label and improve the model performance.

3.4 Performance Analysis on Hyper-Parameters

To further validate the effectiveness of each component in MMDA, we explore the influence of hyper-parameters on the performance of our model.

The choice of β in L_{IM}: β is a balancing factor that adjusts the contribution of fair diversity-promoting objective L_{div}. The DR classification performance of MMDA with β from 0.1 to 0.5 is shown in **Table 4**. As reported, both accuracy and sensitivity is improved with the increase of β, and MMDA achieves the best performance when β is set to 0.3. However, when we further increase the value of β, the results start to decrease. We consider that the high β value weakens the effect of L_{ce}, which leads the decision boundary to go through the high-density region.

The choice of γ: γ is a balancing factor of the information maximization loss L_{IM} and the cross-entropy loss in the overall loss L_{oa}. In this section, we investigate the effectiveness of this hyper-parameter. The results shown in **Table 5** demonstrate that MMDA achieves the highest effectiveness when γ is set to 0.3. The cross-entropy loss in the overall loss L_{oa} acts as a guide of the target model. If γ is too small, the effect of the pseudo labels is reduced. If γ is too large, the generalization of the target model will be limited. In order to learn more discriminative features in the target domain and enhance the DR diagnosis performance of the model, it is necessary to adjust the best value of γ.

3.5 Visual Analysis of Model Performance

Furthermore, in order to prove the superiority of the MMDA framework for practical applications, the ROC curve, and t-SNE plot are adopted to visualize our model.

ROC curve: The receiver operating characteristic curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. In the ROC curve, the closer the apex of the curve toward the upper left corner, the greater the discriminatory ability of the test. The ROC curve of MMDA for diabetic retinopathy classification is drawn in **Figure 4**, which obtains the area under the ROC curve of 0.94 and is above the diagonal and close to the point in the upper left corner, demonstrating that MMDA has a satisfying prediction performance.

t-SNE plot: t-Distributed Stochastic Neighbor Embedding (t-SNE) is a technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional data. It maps high-dimensional data to two or more dimensions suitable for human observation. In order to validate the effectiveness of MMDA, we perform a t-SNE plot using the target image features extracted by the trained target feature encoding module (f_t). As shown in **Figure 5**, retinal images of non-referable DR and referable DR are well separated, because MMDA can learn discriminative features to detect referable diabetic retinopathy. The relatively clear boundaries in **Figure 5** suggest that it is practical to train a robust prediction model using MMDA in the absence of labeled target data.

4 DISCUSSION

Retinal images are usually used to build an automatic diabetic retinopathy diagnosis system (Gardner et al., 1996; Acharya et al.,

2009; Ram et al., 2010; Gulshan et al., 2016; Lam et al., 2018; Jiang et al., 2019; Preston et al., 2021). However, whether using traditional machine learning methods (Gardner et al., 1996; Acharya et al., 2009; Ram et al., 2010) or deep supervised learning methods (Gulshan et al., 2016; Lam et al., 2018; Jiang et al., 2019; Preston et al., 2021), they all need a large amount of labeled data during training. In the biomedical image analysis field, labeling work is expensive and the privacy issue is highly sensitive. To tackle this challenge, we consider developing an unsupervised method that the DR diagnosis performance is excellent but labeled retinal images are unnecessary.

In this article, we present a novel MMDA that incorporates model weight mechanism into the MMDA technique. MMDA can be trained in an end-to-end manner with merely unlabeled target retinal images for DR classification. To the best of our knowledge, MMDA is the first attempt to automatically diagnose diabetic retinopathy by adopting an unsupervised domain adaptation technique with multiple source models. The main advantage of this article is that the MMDA can learn helpful knowledge only from source models without any source data, which can relieve the limitation of data privacy from different medical agencies.

Our proposed MMDA method aims to exploit the source knowledge and relationship between the source models and the target model, instead of learning from labeled retinal images directly, thus helping protect the patients' privacy and no need to label images.

In order to fully explore the discrepancy between each source domain and target domain, we propose a model weight mechanism. By incorporating the mechanism into the source classifiers and feature-level clustering-based pseudo-labeling process, the diagnosis performance of the target model is improved.

Extensive experiments and ablation studies on the APTOS2019 dataset demonstrate that MMDA achieves competitive DR diagnosis performance in comparison with state-of-the-art supervised learning methods. However, the DR classification performance still has a distance from the advanced supervised methods due to the discrepancy between source and target models, especially for the invalid access to source data.

Model visualizations (**Figures 4**, **5**) suggest that non-referable (grade 0, 1) and referable cases (from grade 2 to 4) can be diagnosed well. We will focus on the fine-grained classification of the DR grading task (Zheng et al., 2017) in the future.

5 CONCLUSION

When incorporating deep learning techniques in the automatic DR diagnosis system, time-consuming labeling work and privacy issues are critical problems. The present study is designed to exploit existing models and unlabeled retinal images for DR diagnosis to resolve these issues. Ablation studies show the effectiveness of our proposed modules, and the comparison with state-of-the-art supervised learning approaches demonstrates the superiority of our method. Moreover, model visualization indicates that our method can effectively diagnose non-referable and referable cases, with excellent diagnosing results.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

GZ, BS, and ZZ contributed to conception and design of the study. JP organized the database and performed the statistical analysis. GZ and BS wrote the first draft of the manuscript. ZZ wrote the revised manuscript. WY and YL wrote sections of the manuscript, and supervised this work. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the Research Funds of Shanxi Transformation and Comprehensive Reform Demonstration Zone (Grant No. 2018KJCX04), the Fund for Shanxi "1331 Project," and the Key Research and Development Program of Shanxi Province (Grant No. 201903D311009). The work was also partially sponsored by the Research Foundation of the Education Bureau of Shanxi Province (Grant No. HLW-20132), the Scientific Innovation Plan of Universities in Shanxi Province (Grant No. 2021L575), the Shanxi Scholarship Council of China (Grant No. 2020-149), the Innovation and entrepreneurship training program for College Students (NO.202111242006X), the Medical Science and Technology Development Project Fund of Nanjing (Grant No. YKK21262), and the Nanjing Enterprise Expert Team Project. This work was also supported by the Medical Big Data Clinical Research Project of Nanjing Medical University, Medical and Health Research Project of Zhejiang Province (2018PY066).

REFERENCES

Acharya, U. R., Lim, C. M., Ng, E. Y. K., Chee, C., and Tamura, T. (2009). Computer-based Detection of Diabetes Retinopathy Stages Using Digital Fundus Images. *Proc. Inst. Mech. Eng. H.* 223, 545–553. doi:10.1243/09544119JEIM486

Ai, Z., Huang, X., Fan, Y., Feng, J., Zeng, F., and Lu, Y. (2021). DR-IIXRN : Detection Algorithm of Diabetic Retinopathy Based on Deep Ensemble Learning and Attention Mechanism. *Front. Neuroinform* 15, 778552. doi:10.3389/fninf.2021.778552

Antonetti, D. A., Silva, P. S., and Stitt, A. W. (2021). Current Understanding of the Molecular and Cellular Pathology of Diabetic Retinopathy. *Nat. Rev. Endocrinol.* 17, 195–206. doi:10.1038/s41574-020-00451-4

Bodapati, J. D., Shaik, N. S., and Naralasetti, V. (2021). Deep Convolution Feature Aggregation: an Application to Diabetic Retinopathy Severity Level Prediction. *SIViP* 15, 923–930. doi:10.1007/s11760-020-01816-y

Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., and Ohlrogge, A. W. (2018). Idf Diabetes Atlas: Global Estimates of Diabetes Prevalence for 2017 and Projections for 2045. *Diabetes Res. Clin. Pract.* 138, 271–281. doi:10.1016/j.diabres.2018.02.023

Chou, C.-F., Sherrod, C. E., Zhang, X., Barker, L. E., Bullard, K. M., Crews, J. E., et al. (2014). Barriers to Eye Care Among People Aged 40 Years and Older with Diagnosed Diabetes, 2006–2010. *Diabetes care* 37, 180–188. doi:10.2337/dc13-1507

Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., et al. (2021). A Deep Learning System for Detecting Diabetic Retinopathy across the Disease Spectrum. *Nat. Commun.* 12, 1–11. doi:10.1155/2021/9943067

Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., et al. (2014). Feedback on a Publicly Distributed Image Database: the Messidor Database. *Image Analysis Stereology* 33, 231–234. doi:10.5566/ias.1155

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: A Large-Scale Hierarchical Image Database," in 2009 IEEE conference on computer vision and pattern recognition (IEEE), 248–255. doi:10.1109/cvpr.2009.5206848

Diabetic retinopathy detection (2015). *Diabetic Retinopathy Detection*.

Farag, M. M., Fouad, M., and Abdel-Hamid, A. T. (2022). Automatic Severity Classification of Diabetic Retinopathy Based on Densenet and Convolutional Block Attention Module. *IEEE Access*. doi:10.1109/access.2022.3165193

Ferris, F. L. (1993). How Effective Are Treatments for Diabetic Retinopathy? *Jama* 269, 1290–1291. doi:10.1001/jama.269.10.1290

Gardner, G. G., Keating, D., Williamson, T. H., and Elliott, A. T. (1996). Automatic Detection of Diabetic Retinopathy Using an Artificial Neural Network: a Screening Tool. *Br. J. Ophthalmol.* 80, 940–944. doi:10.1136/bjo.80.11.940

Graham, B. (2015). *Kaggle Diabetic Retinopathy Detection Competition Report*. Coventry: University of Warwick.

Grandvalet, Y., and Bengio, Y., (2005). Semi-supervised Learning by Entropy Minimization. *CAP*, 281–296.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama* 316, 2402–2410. doi:10.1001/jama.2016.17216

He, A., Li, T., Li, N., Wang, K., and Fu, H. (2020). Cabnet: Category Attention Block for Imbalanced Diabetic Retinopathy Grading. *IEEE Trans. Med. Imaging*. 40, 143–153. doi:10.1109/TMI.2020.3023463

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778. doi:10.1109/cvpr.2016.90

Hu, W., Miyato, T., Tokui, S., Matsumoto, E., and Sugiyama, M. (2017). Learning Discrete Representations via Information Maximizing Self-Augmented Training*Int. Conf. Mach. Learn.*, 1558–1567.

Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., and Qian, W. (2019). "An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 2045–2048. doi:10.1109/embc.2019.8857160

Khalifa, N. E. M., Loey, M., Taha, M. H. N., and Mohamed, H. N. E. T. (2019). Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection. *Acta Inform. Medica* 27, 327. doi:10.5455/aim.2019.27.327-332

Kingma, D. P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*

Lam, C., Yi, D., Guo, M., and Lindsey, T. (20182018). Automated Detection of Diabetic Retinopathy Using Deep Learning. *AMIA summits Transl. Sci. Proc.*, 147.

Müller, R., Kornblith, S., and Hinton, G. (2019). When Does Label Smoothing Help? *arXiv preprint arXiv:1906.02629*

MacLennan, P. A., McGwin, G., Searcey, K., and Owsley, C. (2014). A Survey of alabama Eye Care Providers in 2010–2011. *BMC Ophthalmol.* 14, 1–10. doi:10.1186/1471-2415-14-44

Meng, Y., Zhang, H., Gao, D., Zhao, Y., Yang, X., Qian, X., et al. (2021a). "Bi-gcn: Boundary-Aware Input-dependent Graph Convolution Network for Biomedical Image Segmentation," in The British Machine Vision Conference. doi:10.48550/arXiv.2110.14775

Meng, Y., Zhang, H., Zhao, Y., Yang, X., Qiao, Y., MacCormick, I. J., et al. (2021b). Graph-based Region and Boundary Aggregation for Biomedical Image Segmentation. *IEEE Trans. Med. Imaging* 41, 690–701. doi:10.1109/TMI.2021.3123567

Messidor-2 dr grades (2018). *Messidor-2 Dr Grades*.

Mohamed, E., Abd Elmohsen, M., and Basha, T. (2021). "Improved Automatic Grading of Diabetic Retinopathy Using Deep Learning and Principal Component Analysis," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE), 3898–3901. doi:10.1109/embc46164.2021.9630919

Mohamed, Q., Gillies, M. C., and Wong, T. Y. (2007). Management of Diabetic Retinopathy: a Systematic Review. *Jama* 298, 902–916. doi:10.1001/jama.298.8.902

Narayanan, B. N., Hardie, R. C., De Silva, M. S., and Kueterman, N. K. (2020). Hybrid Machine Learning Architecture for Automated Detection and Grading of Retinal Images for Diabetic Retinopathy. *J. Med. Imaging* 7, 034501. doi:10.1117/1.JMI.7.3.034501

Owsley, C., McGwin, G., Scilley, K., Girkin, C. A., Phillips, J. M., and Searcey, K. (2006). Perceived Barriers to Care and Attitudes about Vision and Eye Care: Focus Groups with Older African Americans and Eye Care Providers. *Investigative Ophthalmol. Vis. Sci.* 47, 2797–2802. doi:10.1167/iovs.06-0107

Porwal, P., Kokare, R., and Pachade, S. (2018). Indian Diabetic Retinopathy Image Dataset (Idrid). *Data* 18, 25. doi:10.3390/data3030025

Preston, F. G., Meng, Y., Burgess, J., Ferdousi, M., Azmi, S., Petropoulos, I. N., et al. (2021). Artificial Intelligence Utilising Corneal Confocal Microscopy for the Diagnosis of Peripheral Neuropathy in Diabetes Mellitus and Prediabetes. *Diabetologia*, 1–10. doi:10.1007/s00125-021-05617-x

Ram, K., Joshi, G. D., and Sivaswamy, J. (2010). A Successive Clutter-Rejection-Based Approach for Early Detection of Diabetic Retinopathy. *IEEE Trans. Biomed. Eng.* 58, 664–673. doi:10.1109/TBME.2010.2096223

Roychowdhury, S., Koozekanani, D. D., and Parhi, K. K. (2013). Dream: Diabetic Retinopathy Analysis Using Machine Learning. *IEEE J. Biomed. health Inf.* 18, 1717–1728. doi:10.1109/JBHI.2013.2294635

Vives-Boix, V., and Ruiz-Fernández, D. (2021). Diabetic Retinopathy Detection through Convolutional Neural Networks with Synaptic Metaplasticity. *Comput. Methods Programs Biomed.* 206, 106094. doi:10.1016/j.cmpb.2021.106094

Vo, H. H., and Verma, A. (2016). "New Deep Neural Nets for Fine-Grained Diabetic Retinopathy Recognition on Hybrid Color Space," in 2016 IEEE International Symposium on Multimedia (ISM) (IEEE), 209–215. doi:10.1109/ism.2016.0049

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). "Aggregated Residual Transformations for Deep Neural Networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 1492–1500. doi:10.1109/cvpr.2017.634

Yoo, T. K., and Park, E.-C. (2013). Diabetic Retinopathy Risk Prediction for Fundus Examination Using Sparse Learning: a Cross-Sectional Study. *BMC Med. Inf. Decis. Mak.* 13, 1–14.

Zheng, H., Fu, J., Mei, T., and Luo, J. (2017). "Learning Multi-Attention Convolutional Neural Network for Fine-Grained Image Recognition," in Proceedings of the IEEE international conference on computer vision, 5209–5217. doi:10.1109/iccv.2017.557

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, Zhang, Pan, Yang 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 Genetics

ORIGINAL RESEARCH
published: 08 July 2022
doi: 10.3389/fgene.2022.896805

A Predictive Model for the 10-year Overall Survival Status of Patients With Distant Metastases From Differentiated Thyroid Cancer Using XGBoost Algorithm-A Population-Based Analysis

Shuai Jin[1†], Xing Yang[2†], Quliang Zhong[3†], Xiangmei Liu[4], Tao Zheng[1], Lingyan Zhu[5†] and Jingyuan Yang[6*†]*

[1]School of Big Health, Guizhou Medical University, Guiyang, China, [2]School of Medicine and Health Administration, Guizhou Medical University, Guiyang, China, [3]Department of Urology, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [4]School of Clinical Medicine, Guizhou Medical University, Guiyang, China, [5]Health Management Center, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [6]School of Public Health, Guizhou Medical University, Guiyang, China

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Eman Toraih,
Tulane University, United States
Lei Zhu,
Fifth Affiliated Hospital of Wenzhou Medical University, China
Hongzhou Liu,
Chinese PLA General Hospital, China

***Correspondence:**
Lingyan Zhu
zhuling_happy@126.com
Jingyuan Yang
yangjingyuan@gmc.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *15 March 2022*
Accepted: *16 June 2022*
Published: *08 July 2022*

Citation:
Jin S, Yang X, Zhong Q, Liu X, Zheng T, Zhu L and Yang J (2022) A Predictive Model for the 10-year Overall Survival Status of Patients With Distant Metastases From Differentiated Thyroid Cancer Using XGBoost Algorithm-A Population-Based Analysis. Front. Genet. 13:896805. doi: 10.3389/fgene.2022.896805

Purpose: To explore clinical and non-clinical characteristics affecting the prognosis of patients with differentiated thyroid cancer with distant metastasis (DTCDM) and establish an accurate overall survival (OS) prognostic model.

Patients and methods: Study subjects and related information were obtained from the National Cancer Institute's surveillance, epidemiology, and results database (SEER). Kaplan-Meier analysis, log-rank test, and univariate and multivariate Cox analysis were used to screen for factors influencing the OS of patients with DTCDM. Nine variables were introduced to build a machine learning (ML) model, receiver operating characteristic (ROC) was used to evaluate the recognition ability of the model, calibration plots were used to obtain prediction accuracy, and decision curve analysis (DCA) was used to estimate clinical benefit.

Results: After applying the inclusion and exclusion criteria, a total of 3,060 patients with DTCDM were included in the survival analysis from 2004 to 2017. A machine learning prediction model was developed with nine variables: age at diagnosis, gender, race, tumor size, histology, regional lymph node metastasis, primary site surgery, radiotherapy, and chemotherapy. After excluding patients who survived <120 months, variables were sub-coded and machine learning was used to model OS prognosis in patients with DTCDM. Patients 6–50 years of age had the highest scores in the model. Other variables with high scores included small tumor size, male sex, and age 51–76. The AUC and calibration curves confirm that the XGBoost model has good performance. DCA shows that our model can be used to support clinical decision-making in a 10-years overall survival model.

Conclusion: An artificial intelligence model was constructed using the XGBoost algorithms to predict the 10-years overall survival rate of patients with DTCDM. After

model validation and evaluation, the model had good discriminative ability and high clinical value. This model could serve as a clinical tool to help inform treatment decisions for patients with DTCDM.

Keywords: differentiated thyroid cancer, Xgboost algorithm, machine learning, distant metastases, predictive model, SEER database

INTRODUCTION

Differentiated thyroid cancer (DTC) is the most common endocrine malignancy, with the global incidence increasing dramatically in recent decades (Cabanillas et al., 2016; Siegel et al., 2020). Most DTC patients have a good long-term prognosis because of biological characteristics and effective responses to treatment modalities (Lim et al., 2017; Liu et al., 2018). For TC patients with distant metastasis (DM), however, the overall prognosis is significantly worse (Durante et al., 2006; Sugitani et al., 2008; Farooki et al., 2012; Nies et al., 2021).

The main histological subtypes of DTC include papillary thyroid carcinoma (PTC), and follicular thyroid carcinoma (FTC), <10% of people with DTC will develop DM (Durante et al., 2006). Most DTCDM is asymptomatic and only detected during systematic surveillance or systemic metastatic examination of malignant lymph nodes. The common site of distant metastases from thyroid cancer is the lung, followed by the bone, brain, and liver (Chesover et al., 2021; Liu et al., 2021). Because the incidence of DM is low and there is often an absence of symptoms, it is frequently overlooked at the time of initial TC diagnosis. Ten-year survival rates are often used to assess DTC treatment efficacy and characterize risk factors. For patients with metastatic thyroid cancer, treatment is often individualized and usually consists of thyroid surgery with adjuvant radiotherapy or chemotherapy (Sampson et al., 2007). While prognostic models have been developed to study factors influencing different primary cancer subtypes (Zhao et al., 2019; Zhou et al., 2020; Jin et al., 2021; Kong et al., 2021), the 10-years overall survival of patients with DTCDM is unclear, especially because OS has not

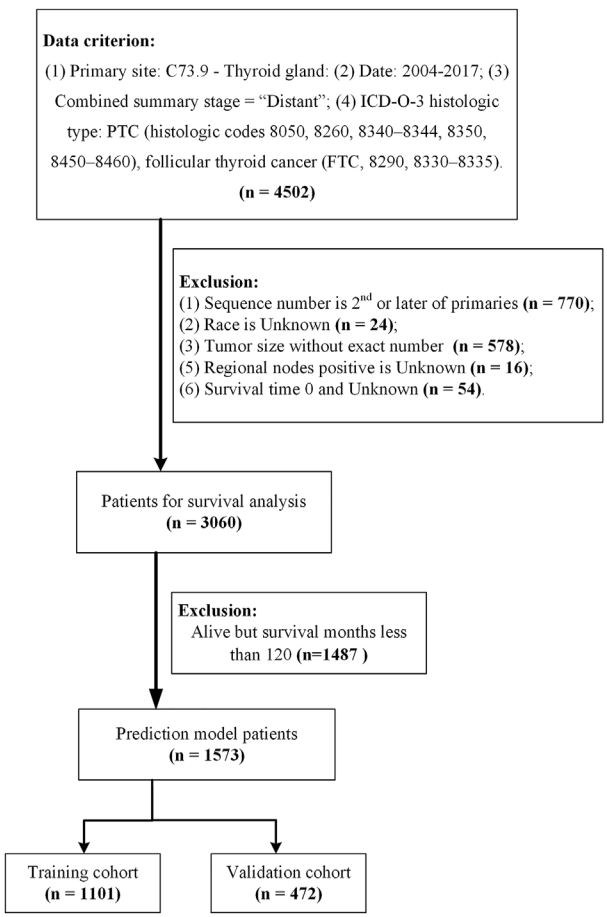


FIGURE 1 | Sample screening process.

changed significantly in recent decades (Goffredo et al., 2013). In addition, there are few prognostic models for DTCDM that can help to inform patient follow-up and treatment decisions (Chen et al., 2021).

Machine learning (ML) is an emerging multi-disciplinary approach used to correlate multiple discrete variables and accurately predict outcomes. Following the development of evidence-based medicine and the need for more advanced tools to collect medical data with complex structures and large sample sizes, ML emerged as an alternative approach to disease diagnosis and prognosis, with high predictive performance and a wide range of applications (Goecks et al., 2020; May, 2021). ML algorithms are now being successfully applied to predict cancer survival (Angraal et al., 2020). The XGBoost algorithm (XGB), in particular, is shown to have excellent prediction performance in previous studies (Senders et al., 2020; Jiang et al., 2021).

The Surveillance, Epidemiology, and End Results (SEER) program is a population-based cancer registry system sponsored by the United States' National Cancer Institute (NCI) that currently covers about 28% of the population in 18 registered states (Noone et al., 2016). Using the considerations listed above, a prognostic modeling analysis of patients with DTCDM was conducted using SEER data. This study assessed the ability of clinical and non-clinical factors to predict DTCDM using the XGB model. The XGB model was also built to predict the 10-years OS rate of TC patients with DM. The performance of the XGBoost model using logistic regression (LR), random forests (RF), and support vector machine (SVM) models were compared.

MATERIALS AND METHODS

Study Population

This study used the SEER database (https://seer.cancer.gov/) developed by the National Cancer Institute, a free cancer registry in the United States. A data use agreement was signed with the SEER database official and all authors followed the specified conditions. Because SEER data is freely available to researchers and patient personal information is officially withheld, no moral or ethical support from the host institution was required for this study.

Data was downloaded from the SEER 18 regs plus database using SEER*Stat (Version 8.3.9.2, https://seer.cancer.gov/data-software/). Data were selected from patients with histologically confirmed distant metastatic thyroid cancer using the following criteria: 1) primary site code is C73.9 - thyroid gland, 2) dates ranging from 2004 to 2017, 3) PTC (histologic codes 8050, 8260, 8340–8344, 8350, 8450–8460), FTC (8290, 8330–8335), and 4) diagnosis confirmation combined with a summary stage for the distant future. Exclusion criteria included 1) sequence numbers for second or later occurrences, 2) unknown race, 3) unclear tumor size, 4) unknown surgery, 5) unknown regional nodes positive (RN_positive), and 6) survival time of 0 or unknown. A total of 3,060 patients were included in the survival analysis to clarify possible factors influencing the prognostic model. Of these, 1,487 patients who had survived <120 months by the follow-up cut-off date were excluded and 1,573 patients were included in the prediction model and grouped in the training set (n = 1,101; 70%) or the validation set (n = 472; 30%) at a 7:3 ratio (**Figure 1**).

Variable Selection and Endpoints

To take full advantage of ML, several established demographic and clinical characteristics were selected as independent variables for analysis. Pathology variables commonly used in thyroid cancer research, including tumor size, histological type, regional lymph node status, surgical modality, radiotherapy, and chemotherapy, and demographic indicators, including age, gender, and ethnicity were included. Age and tumor size were optimally stratified for processing using X-tile software (https://medicine.yale.edu/lab/rimm/research/software/) and included in the survival analysis. These analyses were performed before excluding patients who survived for <120 months. The 10-years OS rate for patients with DTCDM was defined as the model endpoint.

Statistical Analysis

All statistical analyses and model building in this study were performed using R (version 4.1.2, https://www.r-project.org/). The Kaplan–Meier method, with both univariate and multifactorial Cox, was used to screen for OS prognostic factors in thyroid cancer patients, and all variables were used to construct prognostic models. The Chi-square test was then used to analyze differences between the training and validation cohorts. The training set was used to build the XGBoost model and the model was evaluated with the test set. The model capability evaluation includes the following three items: 1) receiver operating characteristic (ROC) curves were used to analyze model discrimination and the area under the ROC curve (AUC) was used to assess predictive model accuracy (Hanley and McNeil, 1982; Wolbers et al., 2009); 2) calibration plots were used to assess the performance of the model, which calibrates how well model predictions agree with the actual observations (Leonard et al., 2020); and 3) decision curve analysis (DCA) was used to assess the clinical usefulness and net benefit of predictive model performance by calculating the difference between the true and false positive rates, weighted by the probability of the chosen risk threshold to assess the net benefit of the model (Vickers and Elkin, 2006). Logistic regression, SVM, and random forest models were built for comparison to the XGBoost model.

RESULTS

Baseline Characteristics and Survival Analysis

A total of 3,060 patients with DTCDM were included in the survival analysis for variable screening. The best cut-off values for age were 50 and 76 years old, and the tumor size was 27 and 65 mm (**Supplementary Figure S1**). Most of the study population (54.02%) was 51–76 years of age with a higher proportion of females than males. Most patients' tumor size was <6.5 cm and the most common histological type was PTC

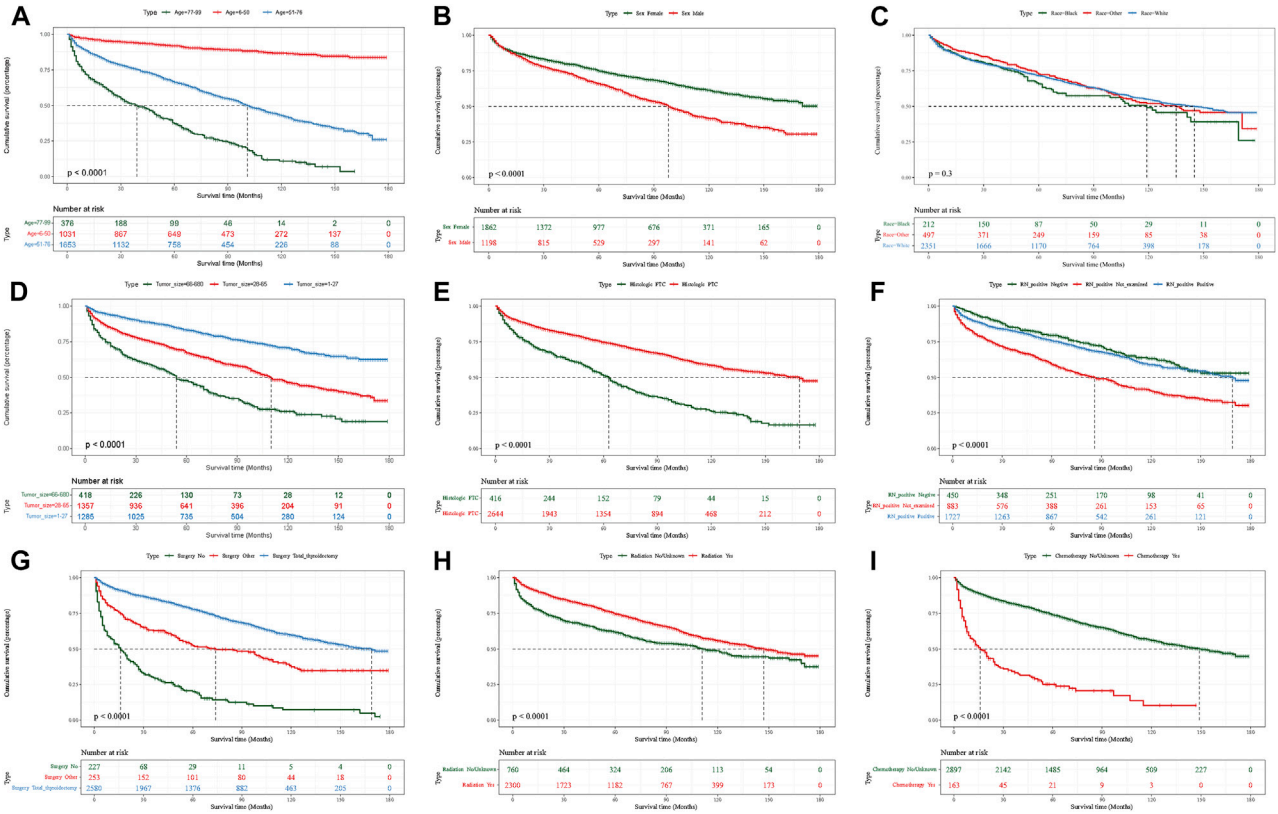


FIGURE 2 | Kaplan–Meier survival curves to evaluate the impact of each classified characteristic.

(86.41%), and FTC type has about 14%. Above half of patients had regional node metastases (56.44%) and total thyroidectomy, radiotherapy, and chemotherapy were used for 56.44%, 76.16%, and 5.33% of patients, respectively. After validation using the Kaplan–Meier method and log-rank test, only survival rates of DTCDM patients did not differ between races (**Figure 2C**), and all other variables differed at all scales (**Figures 2A,B,D–I**). We performed univariate and multivariate Cox models for age, gender, race, tumor size, histological type, RN_positive, surgery, radiotherapy, and chemotherapy, respectively. The multivariate model showed that all variables were independent prognostic factors for DTCDM, except for the race **(Table 1**).

Prognostic Model Construction

Using survival analysis, the following nine variables were included in the prognostic model: age, sex, race, tumor size, histological type, RN_positive, surgery, radiation, and chemotherapy. Most patients (65.9%) with DTCDM ≥10 years were 51–76 years of age, and a higher proportion were women (59.0%) and had PTC histologic staging (81.8%), total thyroidectomy (76.7%), radiotherapy (72.3%), and chemotherapy (7.6%). Analysis of the differences between the two groups for each variable showed no statistical difference, indicating good comparability between the training and validation sets **(Table 2**).

Unlike associated studies (Hou et al., 2020; Wei et al., 2021), categorical variables were included in the prediction model as "dummy variables", age: >76 years, sex: female, race: black, histological type: FTC, tumor size: >65 mm, surgery: no, radiation: no, chemotherapy: no, and RN_positive negative as a control. The XGB algorithm identifies the importance of features based on the magnitude of the gain value obtained for each variable (relative importance scores out of 100), with higher values indicating greater importance to the predicted target. The variables with the highest scores are: age: 6–50 years (39points), tumor size: 1–27 mm (11 points), sex: male (7 points), age: 51–76 years (6 points), RN_positive: not examined (5 points), tumor size: 28–65 mm (5 points), and radiation: yes (5 points) **(Figure 3)**. These variables were included in the LR, SVM, RF models, along with the XGB model for performance testing.

Model Performance

The optimal model performance parameters were determined after several validations and debugging. To assess the ability of the XGB, SVM, LR, and RF models to identify the OS of patients with 10-years DTCDM, the training and validation sets were plotted, and the AUC was calculated.

In the training cohort, the AUC of the XGB model was 0.948, higher than the AUC of the other models (SVM AUC, 0.888; LR AUC 0.873, RF AUC,0.881), and The AUC of XGB in the test group was 0.864, slightly lower than the AUCs of LR and SVM (0.889 and 0.871) and higher than the AUC of the RF model (0.858) **(Figure 4).** The calibration plots of the 10-years TCDM OS for the training and

TABLE 1 | The baseline characteristics, univariate and multivariate cox analysis.

Characteristics	Total	Univariate		Multivariate	
	(*n*, %)	HR (95% CI)	*p*	HR (95% CI)	*p*
Age at diagnosis, years					
77–99	376 (12.29)	References		References	
6–50	1031 (33.69)	0.07 (0.06–0.09)	<0.001	0.10 (0.08–0.13)	<0.001
51–76	1653 (54.02)	0.41 (0.36–0.47)	<0.001	0.49 (0.43–0.57)	<0.001
Sex					
Female	1862 (60.85)	References		References	
Male	1198 (39.15)	1.59 (1.41–1.79)	<0.001	1.27 (1.12–1.43)	<0.001
Race					
Black	212 (6.93)	References		References	
Other	497 (16.24)	0.82 (0.64–1.07)	0.141	0.98 (0.75–1.27)	0.858
White	2351 (76.83)	0.85 (0.68–1.06)	0.143	1.04 (0.83–1.31)	0.714
Tumor size, mm					
66–680	418 (13.66)	References		References	
28–65	1357 (44.35)	0.53 (0.46–0.61)	<0.001	0.7 (0.70–0.82)	<0.001
1–27	1285 (41.99)	0.25 (0.21–0.29)	<0.001	0.41 (0.34–0.49)	<0.001
Histologic					
FTC	416 (13.59)	References		References	
PTC	2644 (86.41)	0.42 (0.36–0.48)	<0.001	0.78 (0.66–0.91)	<0.001
RN positive					
Negative	450 (14.71)	References		References	
Not examined	883 (28.86)	2.08 (1.71–2.52)	<0.001	1.34 (1.09–1.64)	0.005
Positive	1727 (56.44)	1.15 (0.95–1.39)	0.145	1.47 (1.21–1.79)	<0.001
Surgery					
No	227 (7.42)	References		References	
Other	253 (8.27)	0.33 (0.26–0.41)	<0.001	0.53 (0.42–0.67)	<0.001
Total thyroidectomy	2580 (84.31)	0.16 (0.13–0.18)	<0.001	0.35 (0.28–0.42)	<0.001
Radiation					
No	760 (24.84)	References		References	
Yes	2300 (76.16)	0.66 (0.58–0.75)	<0.001	0.86 (0.75–0.98)	0.029
Chemotherapy					
No	2897 (94.67)	References		References	
Yes	163 (5.33)	4.73 (3.89–5.75)	<0.001	2.55 (2.07–3.13)	<0.001

RN, positive, regional nodes metastases.

test sets showed good linear agreement between predicted and actual observations from the XGB model. The XGB and LR models fit better than the SVM and RF models **(Figure 5)**. The DCA curves for the XGB, SVM, RF, and LR models were plotted for the training and test cohorts. The *y*-axis of the decision curve represents the net benefit, the decision analysis metric that determines whether the benefits of a particular clinical decision outweigh the harms. Each point on the *x*-axis represents a threshold probability that distinguishes between patients who die and those who survive. The analysis shows that the XGB, LR, RF, and SVM models all achieve a net clinical benefit **(Figure 6)**.

DISCUSSION

Survival prediction is very difficult in malignancy but important for treatment planning and patient management (Cheon et al., 2016). Compared to the empirical predictions of clinicians, the XGBoost model provides a more reliable option for the 10-years survival status of patients with distant metastases from thyroid cancer. The current study found that XGB had a good predictive value and could help clinicians to develop a rational individualized treatment and management plan. Although thyroid cancer is a relatively slow-growing cancer, once distant metastasis has occurred, the tumor grows exponentially at the location of metastasis, explaining why patients with DM have a worse prognosis (Rajan et al., 2020). While DTC generally has a favorable prognosis, the clinical course can be poor (See et al., 2017). Assessing the survival of patients with DTCDM is thus of great clinical importance.

This study used survival analysis to screen for factors that may affect the OS of patients with DTCDM. Although, the multiple Cox model we developed showed that race is not a prognostic factor for DTCDM patients. One study showed that race does affect the survival of patients with differentiated thyroid cancer and that treatment options need to be specific to different racial groups (Tang et al., 2018). However, Crepeau et al. (Crepeau et al., 2021) found that thyroid cancer prognosis and recurrence did not differ by race following the same surgical approach, especially when all patients receive the same quality of care. It is likely that differences in the prognosis of TC patients by race are the result of social and economic differences between racial groups.

The XGB algorithm is a new type of AI algorithm that is easy to use, efficient and accurate. This algorithm is becoming increasingly popular in the medical field and is now widely used for disease prediction and early diagnosis. The current study used the prognostic variables obtained from survival

TABLE 2 | The baseline characteristics of the training and validation sets used in the prognostic model.

Characteristics	Total cohort	Training cohort	Validation cohort	χ^2	*p*
	n (%)	*n* (%)	*n* (%)		
	1573 (100.0)	1101 (70.0)	472 (30.0)		
Age at diagnosis, years				2.011	0.366
77–99	281 (17.9)	202 (18.3)	79 (16.7)		
6–50	375 (23.8)	252 (22.9)	123 (26.1)		
51–76	917 (58.3)	647 (58.8)	270 (57.2)		
Sex				0.003	0.959
Female	928 (59.0)	650 (59.0)	278 (58.9)		
Male	645 (41.0)	451 (41.0)	194 (41.1)		
Race				1.895	0.388
Black	108 (6.9)	81 (7.4)	27 (5.7)		
Other	256 (16.3)	183 (16.6)	73 (15.5)		
White	1209 (76.9)	837 (76.0)	372 (78.8)		
Tumor size, mm				0.752	0.687
66–680	273 (17.4)	197 (17.9)	76 (16.1)		
28–65	745 (47.4)	519 (47.1)	226 (47.9)		
1–27	555 (35.3)	385 (35.0)	170 (36.0)		
Histologic				1.961	0.161
FTC	286 (18.2)	210 (19.1)	76 (16.1)		
PTC	1287 (81.8)	891 (80.9)	396 (83.9)		
RN positive				0.18	0.914
Negative	220 (14.0)	154 (14.0)	66 (14.0)		
Not examined	585 (37.2)	413 (37.5)	172 (36.4)		
Positive	768 (48.8)	534 (48.5)	234 (49.6)		
Surgery				0.944	0.624
No	189 (12.0)	133 (12.1)	56 (11.9)		
Other	178 (11.3)	130 (11.8)	48 (10.2)		
Total thyroidectomy	1206 (76.7)	838 (76.1)	368 (78.0)		
Radiation					
No	436 (27.7)	312 (28.3)	124 (26.3)	0.704	0.401
Yes	1137 (72.3)	789 (71.7)	348 (73.7)		
Chemotherapy				0.042	0.837
No	1453 (92.4)	1018 (92.5)	435 (92.2)		
Yes	120 (7.6)	83 (7.5)	37 (7.8)		

RN, positive, regional nodes metastases.

analysis to develop a prognostic model for DTCDM OS using the XGB algorithm. The model was validated by combining the clinical and non-clinical variables and was shown to be highly effective. Additional factors that may affect OS and provide more clinical information about DTCDM, were also reported. Traditional ML LR also performed well, possibly because of the study data or because LR performs just as well as ML in clinical prediction models (Christodoulou et al., 2019).

Age was the highest-scoring, and thus most significant, variable in the XGB model, Patients <51 years of age had significantly higher OS than patients >80 years of age. The role for age in predicting TC has been confirmed by other studies, with older patients having a poorer prognosis than younger patients (Sampson et al., 2007; Nixon et al., 2012). Compared with the 8th edition of the TNM staging system with a cut-off value of 55 years, the optimal cut-off values for age in this study were 50 and 77 (Amin et al., 2017). This implies that the cut-off value for age may need to be studied in depth for patients with distant metastatic thyroid cancer. In the current study, there was also a significant score for the 51–79 age group. These findings suggest that treatment should be tailored to patients in different age groups.

The results of one study confirm that patients with PTC with distant metastases have a good prognosis after treatment (Sampson et al., 2007). Another study found that sex is a prognostic factor for DTCDM, likely because estrogen production limits cancer progression (Suteau et al., 2021). The current study found that tumor size was an important factor affecting DTCDM OS, with relatively significant scores for both 1–27 mm and 28–65 mm, a finding supported by Nguyen et al. (Nguyen et al., 2018). Han et al. reported that 15–20 mm tumors do not affect the OS of TCDM patients (Han et al., 2017). Unexamined metastases and those localized in the lymph nodes also scored highly. The number of lymph node metastases correlates strongly with the presence of DM while the risk of DM can be assessed based on the number of lymph nodes (Jeon et al., 2016). These findings may help to resolve controversy over the indication of lateral lymph node dissection (Fan et al., 2018). The regional metastatic status of the lymph nodes should be assessed in all patients with DTCDM.

Radiotherapy had a relatively significant score in the OS prognostic model of patients with DTCDM. Studies indicate that radioactive iodine (RAI) treatment is very effective in patients with small metastases, indicating that early diagnosis improves outcomes (Durante et al., 2006; Sampson et al., 2007). Diagnosis of DM and

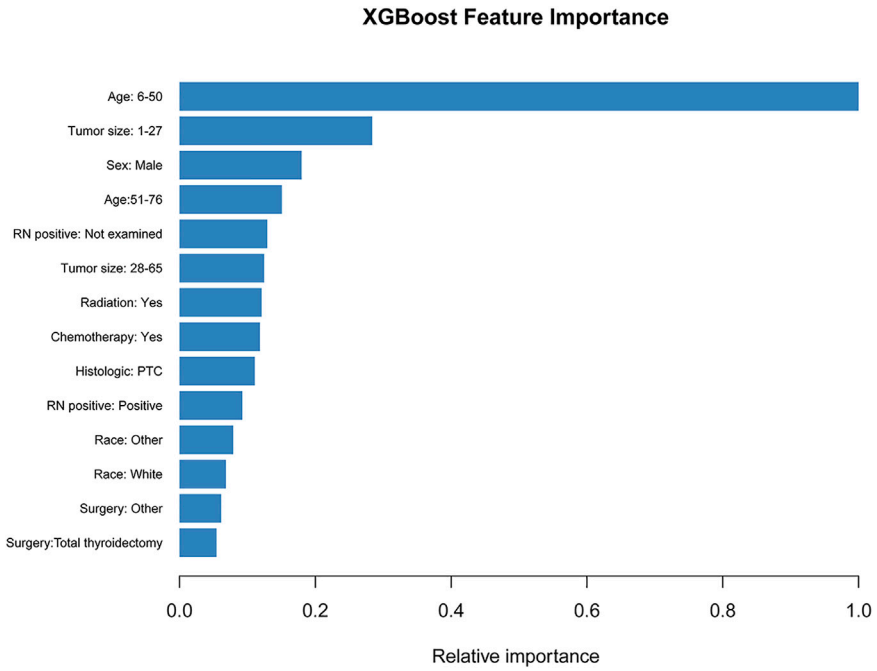


FIGURE 3 | The XGB model was used to calculate the importance of each feature. The bar chart depicts the relative significance of the variables.

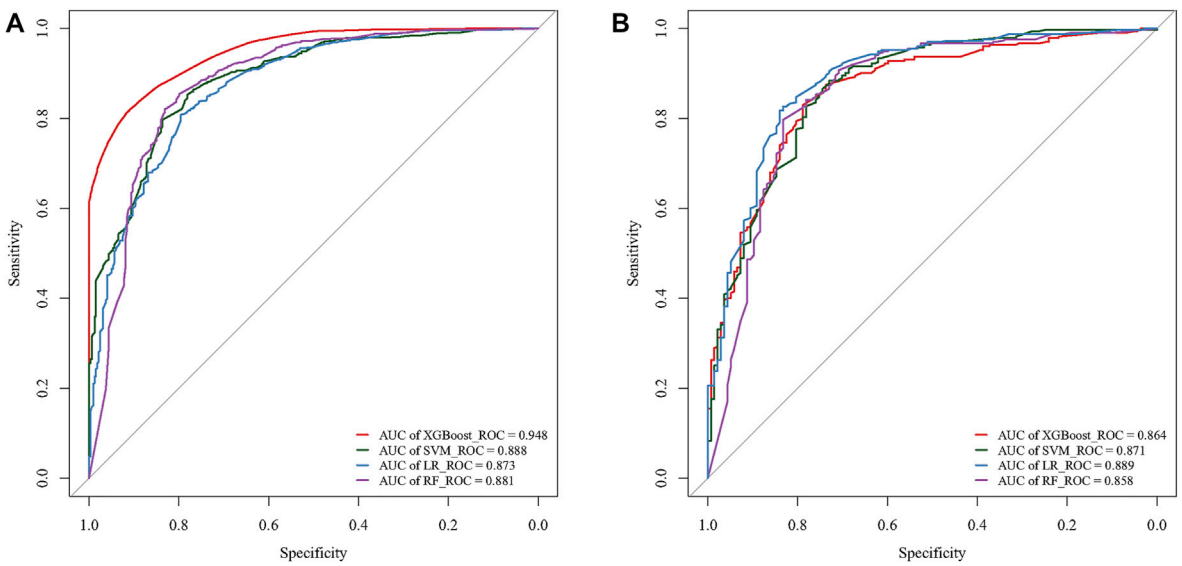


FIGURE 4 | ROC curves of the models for the training **(A)** and test **(B)** cohorts.

initiation of RAI therapy before overt metastases appear, especially in children and adolescents for whom selective treatment is more appropriate (Sugino et al., 2020). While RAI treatment is beneficial for TC survival, however, high-quality RAI accumulation may increase the risk of secondary tumor mutations and more aggressive disease, thus negatively impacting patient survival (Su et al., 2015; Pasqual et al., 2022). Nies et al. concluded that repeated RAI treatment is unlikely to benefit TC patients and may do more harm than good over their lifespan (Nies et al., 2021). In summary, studies differ on whether and how to treat DTC patients with RAI (Jeon et al., 2016; Lin et al., 2018).

Total thyroidectomy and other surgical procedures were important in the prognostic model. Sampson et al. concluded that a total thyroidectomy should be performed alongside RAI treatment (Sampson et al., 2007). It is also suggested that, where

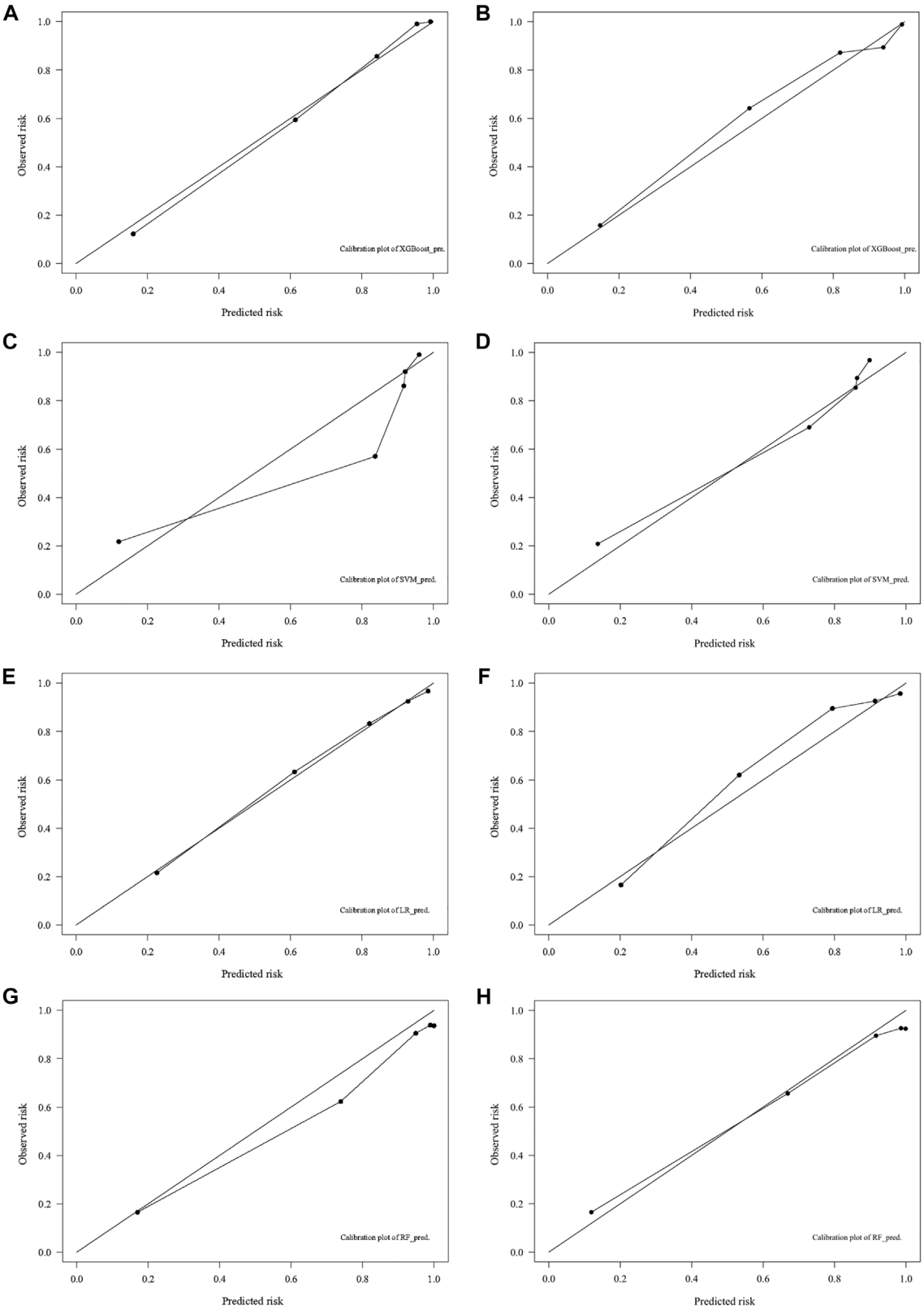


FIGURE 5 | Calibration plots for predicting 10-years DTC development with DM in the training **(A,C,E,G)** and test **(B,D,F,H)** cohorts. DTC: differentiated thyroid cancer, DM: distant metastasis, XGB: XGBoost, SVM: support vector machines, RF: random forest, LR: logistic regression.

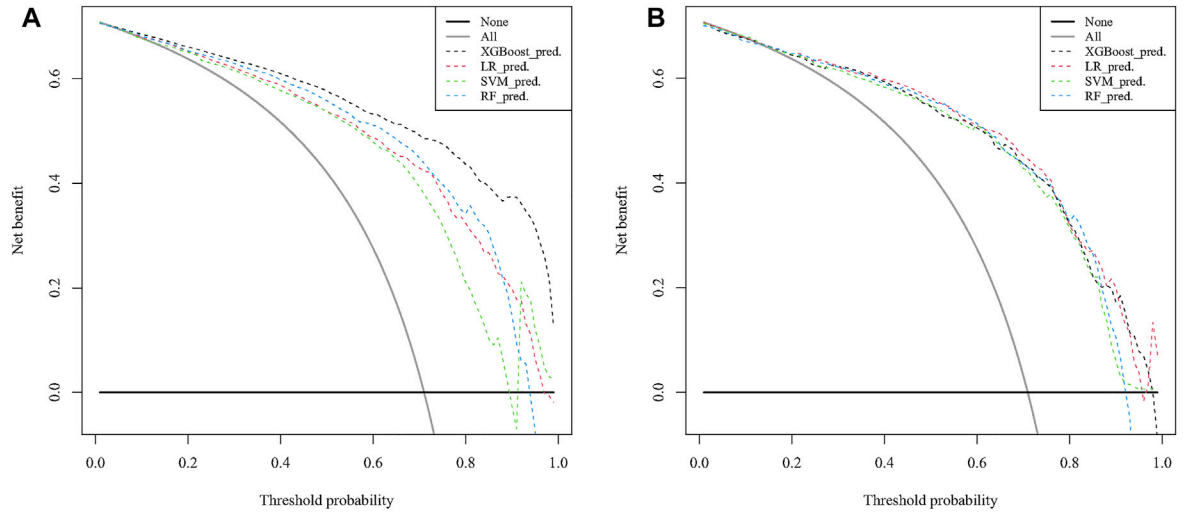


FIGURE 6 | Decision curve analysis to predict 10-years DTC development with DM in the training **(A)** and test **(B)** cohorts. DTC: differentiated thyroid cancer, DM: distant metastasis, XGB: XGBoost, SVM: support vector machines, RF: random forest, LR: logistic regression.

possible, local curative surgery with RAI and thyroid hormone suppression should be performed in patients with DTCDM (Ito et al., 2010). However, the survival benefit of thyroid cancer surgery may vary depending on the site of metastasis (Besic et al., 2016). The current study found that chemotherapy was a strong predictor of OS in patients with DTCDM and a risk factor for OS in survival analysis. Chemotherapy is often administered to patients with large tumors who are no longer candidates for surgery or show iodine resistance. However, a study indicates that chemotherapy is highly toxic and is associated with a poor response rate (Schmidbauer et al., 2017). The high mortality rate of chemotherapy patients may be due to the relative severity of the disease in this group of patients, in addition to the toxic impact of the treatment. Recent studies have shown that targeted therapies such as tyrosine kinase inhibitors (TKIs) offer high survival rates and that patients may have a better outcome if targeted treatments are combined with chemotherapy (Kraeber-Bodéré et al., 2010; Carling and Udelsman, 2014; Viola et al., 2016; Lorusso et al., 2021).

Although the predictive model used in this study had a good performance, there were some limitations. First, the study relied on regression data and some samples with missing information were removed, which may have biased the model. Second, outcome data for individuals receiving targeted therapies were not included in the sample, which may have made the prediction model less comprehensive. Finally, more work needs to be done to explain the predictive efficacy of ML versus traditional statistical methods.

CONCLUSION

This study analyzed the clinical characteristics and prognosis of patients with DTCDM and constructed prognostic models using four machine learning methods. The XGB model was effective at predicting the 10-years OS of patients with DTCDM and may help clinicians to make more accurate and personalized clinical decisions. This is particularly important to improve the long-term prognosis of high-risk 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 in the article/**Supplementary Material**.

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 for participation was not required for this study in accordance with the national legislation and the institutional requirements. Ethical review and approval was not required for the animal study because The SEER database is free for researchers to download and has removed patient information.

AUTHOR CONTRIBUTIONS

SJ, LZ and JY contributed to the conception and design of the study; SJ, QZ and TZ collected and analyzed data; XY, XL, LZ and QZ wrote and revised the manuscript; All authors reviewed and approved the final version of the manuscript.

FUNDING

This study was supported by funds from the Ministry of Education Industry-Academia Cooperation Collaborative Education Project (202101311012, 202102576027).

ACKNOWLEDGMENTS

We are grateful to all the experts who opened the research topic "Medical knowledge-assisted machine learning techniques in individualized medicine" in the journal Frontiers in Genetics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.896805/full#supplementary-material

REFERENCES

Amin, M. B., Greene, F. L., Edge, S. B., Compton, C. C., GershenwaldGershenwald, J. E., Brookland, R. K., et al. and CTR (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge from a Population-Based to a More "personalized" Approach to Cancer Staging. *CA A Cancer J. Clin.* 67, 93–99. doi:10.3322/caac.21388

Angraal, S., Mortazavi, B. J., Gupta, A., Khera, R., Ahmad, T., Desai, N. R., et al. (2020). Machine Learning Prediction of Mortality and Hospitalization in Heart Failure with Preserved Ejection Fraction. *JACC Heart Fail.* 8 (1), 12–21. doi:10.1016/j.jchf.2019.06.013

Besic, N., Schwarzbartl-Pevec, A., Vidergar-Kralj, B., Crnic, T., Gazic, B., and Marolt Music, M. (2016). Treatment and Outcome of 32 Patients with Distant Metastases of Hürthle Cell Thyroid Carcinoma: a Single-Institution Experience. *BMC Cancer* 16, 162. doi:10.1186/s12885-016-2179-3

Cabanillas, M. E., McFadden, D. G., and Durante, C. (2016). Thyroid Cancer. *Lancet* 388 (10061), 2783–2795. doi:10.1016/s0140-6736(16)30172-6

Carling, T., and Udelsman, R. (2014). Thyroid Cancer. *Annu. Rev. Med.* 65, 125–137. doi:10.1146/annurev-med-061512-105739

Chen, S., Jiang, L., Zhang, E., Hu, S., Wang, T., Gao, F., et al. (2021). A Novel Nomogram Based on Machine Learning-Pathomics Signature and Neutrophil to Lymphocyte Ratio for Survival Prediction of Bladder Cancer Patients. *Front. Oncol.* 11, 703033. doi:10.3389/fonc.2021.703033

Cheon, S., Agarwal, A., Popovic, M., Milakovic, M., Lam, M., Fu, W., et al. (2016). The Accuracy of Clinicians' Predictions of Survival in Advanced Cancer: a Review. *Ann. Palliat. Med.* 5 (1), 22–29. doi:10.3978/j.issn.2224-5820.2015.08.04

Chesover, A. D., Vali, R., Hemmati, S. H., and Wasserman, J. D. (2021). Lung Metastasis in Children with Differentiated Thyroid Cancer: Factors Associated with Diagnosis and Outcomes of Therapy. *Thyroid* 31 (1), 50–60. doi:10.1089/thy.2020.0002

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., and Van Calster, B. (2019). A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models. *J. Clin. Epidemiol.* 110, 12–22. doi:10.1016/j.jclinepi.2019.02.004

Crepeau, P. K., Kulkarni, K., Martucci, J., and Lai, V. (2021). Comparing Surgical Thoroughness and Recurrence in Thyroid Cancer Patients across Race/ethnicity. *Surgery* 170 (4), 1099–1104. doi:10.1016/j.surg.2021.05.001

Durante, C., Haddy, N., Baudin, E., Leboulleux, S., Hartl, D., Travagli, J. P., et al. (2006). Long-term Outcome of 444 Patients with Distant Metastases from Papillary and Follicular Thyroid Carcinoma: Benefits and Limits of Radioiodine Therapy. *J. Clin. Endocrinol. Metab.* 91 (8), 2892–2899. doi:10.1210/jc.2005-2838

Fan, W., Xiao, C., and Wu, F. (2018). Analysis of Risk Factors for Cervical Lymph Node Metastases in Patients with Sporadic Medullary Thyroid Carcinoma. *J. Int. Med. Res.* 46 (5), 1982–1989. doi:10.1177/0300060518762684

Farooki, A., Leung, V., Tala, H., and Tuttle, R. M. (2012). Skeletal-related Events Due to Bone Metastases from Differentiated Thyroid Cancer. *J. Clin. Endocrinol. Metabolism* 97 (7), 2433–2439. doi:10.1210/jc.2012-1169

Goecks, J., Jalili, V., Heiser, L. M., and Gray, J. W. (2020). How Machine Learning Will Transform Biomedicine. *Cell.* 181 (1), 92–101. doi:10.1016/j.cell.2020.03.022

Goffredo, P., Sosa, J. A., and Roman, S. A. (2013). Differentiated Thyroid Cancer Presenting with Distant Metastases: a Population Analysis over Two Decades. *World J. Surg.* 37 (7), 1599–1605. doi:10.1007/s00268-013-2006-9

Han, K., Kim, E.-K., and Kwak, J. Y. (2017). 1.5-2 Cm Tumor Size Was Not Associated with Distant Metastasis and Mortality in Small Thyroid Cancer: A Population-Based Study. *Sci. Rep.* 7, 46298. doi:10.1038/srep46298

Hanley, J. A., and McNeil, B. J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143 (1), 29–36. doi:10.1148/radiology.143.1.7063747

Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., et al. (2020). Predicting 30-days Mortality for MIMIC-III Patients with Sepsis-3: a Machine Learning Approach Using XGboost. *J. Transl. Med.* 18 (1), 462. doi:10.1186/s12967-020-02620-5

Ito, Y., Masuoka, H., Fukushima, M., Inoue, H., Kihara, M., Tomoda, C., et al. (2010). Prognosis and Prognostic Factors of Patients with Papillary Carcinoma Showing Distant Metastasis at Surgery (M1 Patients) in Japan. *Endocr. J.* 57 (6), 523–531. doi:10.1507/endocrj.k10e-019

Jeon, M. J., Kim, W. G., Choi, Y. M., Kwon, H., Lee, Y.-M., Sung, T.-Y., et al. (2016). Features Predictive of Distant Metastasis in Papillary Thyroid Microcarcinomas. *Thyroid* 26 (1), 161–168. doi:10.1089/thy.2015.0375

Jiang, J., Pan, H., Li, M., Qian, B., Lin, X., and Fan, S. (2021). Predictive Model for the 5-year Survival Status of Osteosarcoma Patients Based on the SEER Database and XGBoost Algorithm. *Sci. Rep.* 11 (1), 5542. doi:10.1038/s41598-021-85223-4

Jin, M., Megwalu, U. C., and Noel, J. E. (2021). External Beam Radiotherapy for Medullary Thyroid Cancer Following Total or Near-Total Thyroidectomy. *Otolaryngol. Head. Neck Surg.* 164 (1), 97–103. doi:10.1177/0194599820947696

Kong, N., Xu, Q., Zhang, Z., Cui, A., Tan, S., and Bai, N. (2021). Age Influences the Prognosis of Anaplastic Thyroid Cancer Patients. *Front. Endocrinol.* 12, 704596. doi:10.3389/fendo.2021.704596

Kraeber-Bodéré, F., Salaun, P.-Y., Oudoux, A., Goldenberg, D. M., Chatal, J.-F., and Barbet, J. (2010). Pretargeted Radioimmunotherapy in Rapidly Progressing, Metastatic, Medullary Thyroid Cancer. *Cancer* 116 (S4), 1118–1125. doi:10.1002/cncr.24800

Leonard, S. A., Kennedy, C. J., Carmichael, S. L., Lyell, D. J., and Main, E. K. (2020). An Expanded Obstetric Comorbidity Scoring System for Predicting Severe Maternal Morbidity. *Obstet. Gynecol.* 136 (3), 440–449. doi:10.1097/aog.0000000000004022

Lim, H., Devesa, S. S., Sosa, J. A., Check, D., and Kitahara, C. M. (2017). Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013. *Jama* 317 (13), 1338–1348. doi:10.1001/jama.2017.2719

Lin, J.-D., Kuo, S.-F., Huang, B.-Y., Lin, S.-F., and Chen, S.-T. (2018). The Efficacy of Radioactive Iodine for the Treatment of Well-Differentiated Thyroid Cancer with Distant Metastasis. *Nucl. Med. Commun.* 39 (12), 1091–1096. doi:10.1097/MNM.0000000000000897

Liu, W. C., Li, Z. Q., Luo, Z. W., Liao, W. J., Liu, Z. L., and Liu, J. M. (2021). Machine Learning for the Prediction of Bone Metastasis in Patients with Newly Diagnosed Thyroid Cancer. *Cancer Med.* 10 (8), 2802–2811. doi:10.1002/cam4.3776

Liu, Z., Zeng, W., Huang, L., Wang, Z., Wang, M., Zhou, L., et al. (2018). Prognosis of FTC Compared to PTC and FVPTC: Findings Based on SEER Database Using Propensity Score Matching Analysis. *Am. J. Cancer Res.* 8 (8), 1440

Lorusso, L., Cappagli, V., Valerio, L., Giani, C., Viola, D., Puleo, L., et al. (2021). Thyroid Cancers: From Surgery to Current and Future Systemic Therapies through Their Molecular Identities. *Ijms* 22 (6), 3117. doi:10.3390/ijms22063117

May, M. (2021). Eight Ways Machine Learning Is Assisting Medicine. *Nat. Med.* 27 (1), 2–3. doi:10.1038/s41591-020-01197-2

Nguyen, X. V., Roy Choudhury, K., Tessler, F. N., and Hoang, J. K. (2018). Effect of Tumor Size on Risk of Metastatic Disease and Survival for Thyroid Cancer: Implications for Biopsy Guidelines. *Thyroid* 28 (3), 295–300. doi:10.1089/thy.2017.0526

Nies, M., Vassilopoulou-Sellin, R., Bassett, R. L., Yedururi, S., Zafereo, M. E., Cabanillas, M. E., et al. (2021). Distant Metastases from Childhood Differentiated Thyroid Carcinoma: Clinical Course and Mutational

Landscape. *J. Clin. Endocrinol. Metab.* 106 (4), 1683–1697. doi:10.1210/clinem/dgaa935

Nixon, I. J., Whitcher, M. M., Palmer, F. L., Tuttle, R. M., Shaha, A. R., Shah, J. P., et al. (2012). The Impact of Distant Metastases at Presentation on Prognosis in Patients with Differentiated Carcinoma of the Thyroid Gland. *Thyroid* 22 (9), 884–889. doi:10.1089/thy.2011.0535

Noone, A.-M., Lund, J. L., Mariotto, A., Cronin, K., McNeel, T., Deapen, D., et al. (2016). Comparison of SEER Treatment Data with Medicare Claims. *Med. Care* 54 (9), e55–e64. doi:10.1097/mlr.0000000000000073

Pasqual, E., Schonfeld, S., Morton, L. M., Villoing, D., Lee, C., Berrington de Gonzalez, A., et al. (2022). Association between Radioactive Iodine Treatment for Pediatric and Young Adulthood Differentiated Thyroid Cancer and Risk of Second Primary Malignancies. *Jco* 40 (13), 1439–1449. doi:10.1200/jco.21.01841

Rajan, N., Khanal, T., and Ringel, M. D. (2020). Progression and Dormancy in Metastatic Thyroid Cancer: Concepts and Clinical Implications. *Endocrine* 70 (1), 24–35. doi:10.1007/s12020-020-02453-8

Sampson, E., Brierley, J. D., Le, L. W., Rotstein, L., and Tsang, R. W. (2007). Clinical Management and Outcome of Papillary and Follicular (Differentiated) Thyroid Cancer Presenting with Distant Metastasis at Diagnosis. *Cancer* 110 (7), 1451–1456. doi:10.1002/cncr.22956

Schmidbauer, B., Menhart, K., Hellwig, D., and Grosse, J. (2017). Differentiated Thyroid Cancer-Treatment: State of the Art. *Ijms* 18 (6), 1292. doi:10.3390/ijms18061292

See, A., See, N. G., Tan, N. C., Teo, C., Ng, J., Soo, K. C., et al. (2017). Distant Metastasis as the Sole Initial Manifestation of Well-Differentiated Thyroid Carcinoma. *Eur. Arch. Otorhinolaryngol.* 274 (7), 2877–2882. doi:10.1007/s00405-017-4532-9

Senders, J. T., Staples, P., Mehrtash, A., Cote, D. J., Taphoorn, M. J. B., Reardon, D. A., et al. (2020). An Online Calculator for the Prediction of Survival in Glioblastoma Patients Using Classical Statistics and Machine Learning. *Neurosurgery* 86 (2), E184–e192. doi:10.1093/neuros/nyz403

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590

Su, D. H., Chang, S. H., and Chang, T. C. (2015). The Impact of Locoregional Recurrences and Distant Metastases on the Survival of Patients with Papillary Thyroid Carcinoma. *Clin. Endocrinol.* 82 (2), 286–294. doi:10.1111/cen.12511

Sugino, K., Nagahama, M., Kitagawa, W., Ohkuwa, K., Uruno, T., Matsuzu, K., et al. (2020). Distant Metastasis in Pediatric and Adolescent Differentiated Thyroid Cancer: Clinical Outcomes and Risk Factor Analyses. *J. Clin. Endocrinol. Metab.* 105 (11), e3981–e3988. doi:10.1210/clinem/dgaa545

Sugitani, I., Fujimoto, Y., and Yamamoto, N. (2008). Papillary Thyroid Carcinoma with Distant Metastases: Survival Predictors and the Importance of Local Control. *Surgery* 143 (1), 35–42. doi:10.1016/j.surg.2007.06.011

Suteau, V., Munier, M., Briet, C., and Rodien, P. (2021). Sex Bias in Differentiated Thyroid Cancer. *Ijms* 22 (23), 12992. doi:10.3390/ijms222312992

Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Liao, X., et al. (2018). Racial Disparities of Differentiated Thyroid Carcinoma: Clinical Behavior, Treatments, and Long-Term Outcomes. *World J. Surg. Onc* 16 (1), 45. doi:10.1186/s12957-018-1340-7

Vickers, A. J., and Elkin, E. B. (2006). Decision Curve Analysis: a Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* 26 (6), 565–574. doi:10.1177/0272989x06295361

Viola, D., Valerio, L., Molinaro, E., Agate, L., Bottici, V., Biagini, A., et al. (2016). Treatment of Advanced Thyroid Cancer with Targeted Therapies: Ten Years of Experience. *Endocr. Relat. Cancer* 23 (4), R185–R205. doi:10.1530/erc-15-0555

Wei, L., Huang, Y., Chen, Z., Li, J., Huang, G., Qin, X., et al. (2021). A Novel Machine Learning Algorithm Combined with Multivariate Analysis for the Prognosis of Renal Collecting Duct Carcinoma. *Front. Oncol.* 11, 777735. doi:10.3389/fonc.2021.777735

Wolbers, M., Koller, M. T., Witteman, J. C. M., and Steyerberg, E. W. (2009). Prognostic Models with Competing Risks: Methods and Application to Coronary Risk Prediction. *Epidemiology* 20 (4), 555–561. doi:10.1097/EDE.0b013e3181a39056

Zhao, H., Huang, T., and Li, H. (2019). Risk Factors for Skip Metastasis and Lateral Lymph Node Metastasis of Papillary Thyroid Cancer. *Surgery* 166 (1), 55–60. doi:10.1016/j.surg.2019.01.025

Zhou, L., Li, Q., Chen, S., Huang, Y., Wei, W., Zhang, C., et al. (2020). Synergic Effects of Histology Subtype, Lymph Node Metastasis, and Distant Metastasis on Prognosis in Differentiated Thyroid Carcinoma Using the SEER Database. *Gland. Surg.* 9 (4), 907–918. doi:10.21037/gs-20-273

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 Jin, Yang, Zhong, Liu, Zheng, Zhu 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 Molecular Biosciences

TYPE Original Research
PUBLISHED 22 July 2022
DOI 10.3389/fmolb.2022.913602

Check for updates

OPEN ACCESS

EDITED BY
Yong Teng,
Emory University, United States

REVIEWED BY
Abdulrazak Yahya Saleh,
Universiti Malaysia Sarawak, Malaysia
Dong Xu,
University of Missouri, United States

*CORRESPONDENCE
Magbubah Essack,
magbubah.essack@kaust.edu.sa
Xin Gao,
xin.gao@kaust.edu.sa

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 05 April 2022
ACCEPTED 29 June 2022
PUBLISHED 22 July 2022

CITATION
Albaradei S, Albaradei A, Alsaedi A, Uludag M, Thafar MA, Gojobori T, Essack M and Gao X (2022), MetastaSite: Predicting metastasis to different sites using deep learning with gene expression data.
Front. Mol. Biosci. 9:913602.
doi: 10.3389/fmolb.2022.913602

COPYRIGHT
© 2022 Albaradei, Albaradei, Alsaedi, Uludag, Thafar, Gojobori, Essack and Gao. 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.

MetastaSite: Predicting metastasis to different sites using deep learning with gene expression data

Somayah Albaradei[1,2], Abdurhman Albaradei[3], Asim Alsaedi[4,5], Mahmut Uludag[1], Maha A. Thafar[1,6], Takashi Gojobori[1], Magbubah Essack[1]* and Xin Gao[1]*

[1]Computer Electrical and Mathematical Sciences and Engineering Division (CEMSE), Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [2]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, [3]Al-Hada Armed Forces Hospital, Taif, Saudi Arabia, [4]King Saud Bin Abdulaziz University for Health Sciences, Jeddah, Saudi Arabia, [5]King Abdulaziz Medical City, Jeddah, Saudi Arabia, [6]College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

Deep learning has massive potential in predicting phenotype from different omics profiles. However, deep neural networks are viewed as black boxes, providing predictions without explanation. Therefore, the requirements for these models to become interpretable are increasing, especially in the medical field. Here we propose a computational framework that takes the gene expression profile of any primary cancer sample and predicts whether patients' samples are primary (localized) or metastasized to the brain, bone, lung, or liver based on deep learning architecture. Specifically, we first constructed an AutoEncoder framework to learn the non-linear relationship between genes, and then DeepLIFT was applied to calculate genes' importance scores. Next, to mine the top essential genes that can distinguish the primary and metastasized tumors, we iteratively added ten top-ranked genes based upon their importance score to train a DNN model. Then we trained a final multi-class DNN that uses the output from the previous part as an input and predicts whether samples are primary or metastasized to the brain, bone, lung, or liver. The prediction performances ranged from AUC of 0.93–0.82. We further designed the model's workflow to provide a second functionality beyond metastasis site prediction, i.e., to identify the biological functions that the DL model uses to perform the prediction. To our knowledge, this is the first multi-class DNN model developed for the generic prediction of metastasis to various sites.

KEYWORDS

machine learning, deep learning, artificial intelligence, metastasis, metastasis site, gene expression, clinical decision-making

1 Introduction

Precision medicine is a path that could profoundly change and improve medical practices. This idea proposes using genetic data of individual patients to enhance clinical decision-making, and "omics" technologies now provide a means to acquire such patient data, making precision medicine feasible. Clinical decision-making includes diagnosis, prognosis, choosing the most appropriate treatment, etc. One avenue pursued to support clinical decision-making is building classifiers using gene expression profiles that can function as forms of artificial intelligence (AI).

Many machine learning methods, including support vector machines, random forest, and boosting, are among the primary tools currently being used to make biological discoveries from the vast amount of available gene expression data (Libbrecht and Noble, 2015). However, deep learning (DL) is emerging as a more powerful machine learning method (Goodfellow et al., 2016), although the primary DL application domain is image recognition and speech recognition. Nonetheless, DL is showing promise in many other fields of science, especially in precision medicine and genomics data analysis (Grapov et al., 2018; Martorell-Marugán et al., 2019), as DL can extract intricate structures in high-dimensional data (Najafabadi et al., 2015). However, DL is still new in the bioinformatics community; thus, only a few published works show its application to gene expression-based models (Daoud and Mayo, 2019). Furthermore, unlike images or text data, gene expression data has no clear structure that we can exploit in a neural network architecture. Thus, many new architectures are surfacing for metastasis prediction from gene expression data, such as multilayer perceptron architecture (Albaradei et al., 2019; Albaradei et al., 2021a; Albaradei et al., 2021b) , autoencoder architectures (Sharifi-Noghabi et al., et al.; Albaradei et al., 2021c; Fakoor et al., 2013) and Graph deep learning (Xu et al., 2021). Most of these proposed models try to solve a binary classification problem that classifies samples as metastatic or non-metastatic (Albaradei et al., 2021a). However no generic computational framework based on DL that accepts raw gene expression data to predict whether cancer is primary or has spread to various metastasis sites exists.

The main concern of DL used in medical applications is the lack of interpretability. The reason being, DL networks can be viewed as black boxes that form an input layer (wherein we place the gene expression profile of patients) and an output layer (offering predictions without interpretability). Suppose we do not meet this interpretability criterion at a good standard. In that case, physicians will not be able to trust the decision of the neural network, as they need interpretable data to ensure patients' safety. Specifically, they need data about neurons, genes, and related biological processes involved in the prediction and the decision-making process to make informed decisions. Thus, researchers are now attempting to make the DL networks more interpretable.

In this work, we attempted to develop an AI method that could translate into a tool that supports clinical decision-making with regard to identifying metastasis and pinpointing the metastasis site (Figure 1). In this process, we also show the biological functions that the model uses to perform the prediction. That is, current work that interprets DL models identifies the genes that impact the prediction. Here, we propose interpreting the hidden neurons by linking the neurons to the enriched biological functions. In this work, we developed such a DL model. The DL framework takes as input raw gene expression data for a sample and predicts whether it is primary or metastasized to the brain, bone, lung, or liver. In the first phase, we used AutoEncoder (AE) to reduce the dimension of the expression data. Then, we applied DeepLIFT to compute an importance score (i.e., the impact of each input layer neuron on the latent layer neurons) used to rank the genes. Finally, to mine the genes that can distinguish the primary and tumor samples metastasized to different sites, we iteratively fed ten top-ranked genes (based upon the importance score) to the DNN model for training. In the second phase, we trained and evaluated a final multi-class DNN model to make the metastasis site predictions. Here, we also used the DeepLIFT approach to identify the essential neurons that lead to the prediction and the set of genes that activate these critical neurons. Then, we linked these critical genes to Gene Ontology (GO). We also provided analyses using Molecular Signatures Database (MSigDB) and the Disease Gene Network (DisGeNet) to support and increase the biology extracted from the essential neurons' list of genes.

2 Method and materials

2.1 Gene expression datasets

We searched for gene expression datasets in Gene Expression Omnibus (GEO) (Edgar et al., 2002) using the following query: "metastas* AND (bone OR brain OR lung OR liver) AND Homo sapiens" filtered by "Expression profiling by array" in September 2021. We retrieved 837 entries which we sifted through and found microarray gene expression data for primary tumors (breast, colorectal, kidney, liver, lung, pancreatic, and prostate cancer samples), and tumors metastasized from these primary tumors to the bone, brain, lung, or liver. Table 1 provides the GEO accession numbers of the samples used in this study, along with the sample statistics. Similar to the approach used in (Chereda et al., 2019), we used the RMA probe-summary algorithm (Irizarry et al., 2003) to process each dataset, after which they were combined based on the HG-U133A array probe names, and quantile normalization was applied across all datasets. In cases where multiple probes were mapped to one gene, the probe with the highest average value was taken. Finally, we used the integrated datasets for each of the four sites as input for the DL models. However, before we fed the data to the DL model, we used the synthetic minority oversampling technique (SMOTE)

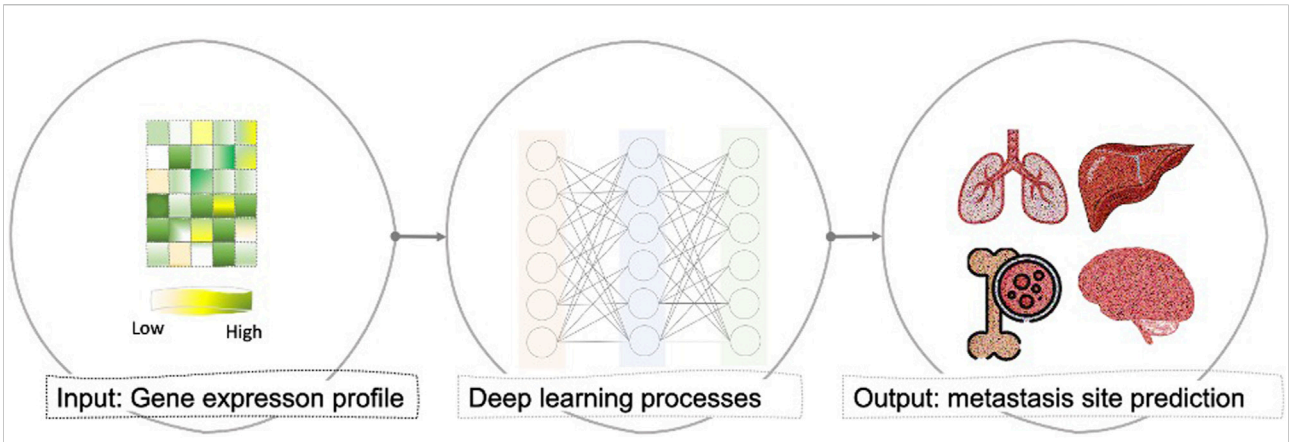


FIGURE 1
General overview of the proposed computational framework that takes the gene expression profile of any primary cancer sample and predicts whether patients' samples are primary (localized) or had been metastasized to brain, bone, lung, or liver based on deep learning architecture.

TABLE 1 The gene expression datasets from GEO with the number of primary and metastasized samples for each site.

	Bone	Brain	Lung	Liver
Breast	220 Primary, 72 Metastasized [GSE 2034, GSE137842]	27 Primary, 65 Metastasized [GSE12276, GSE125989, GSE46928, GSE18549]	47 Primary, 18 Metastasized [GSE16554, GSE5327]	28 Primary, 16 Metastasized [GSE18549]
Colorectal	0	10 Primary, 23 Metastasized [GSE14108]	186 Primary, 47 Metastasized [GSE18549, GSE41258]	219 Primary, 86 Metastasized [GSE41258, GSE18549, GSE6605]
Kidney	0	0	10 Primary, 10 Metastasized [GSE22541]	0
Liver	0	0	31 Primary, 31 Metastasized [GSE141016]	0
Lung	14 Primary, 19 Metastasized [GSE10096]	15 Primary, 23 Metastasized [GSE18549]	0	0
Pancreas	0	0	0	15 Primary, 14 Metastasized [GSE19279]
Prostate	16 Primary, 17 Metastasized [GSE18549, GSE43332]	0	0	0

to oversample the minority class using the imbalanced-learn python library (Chawla et al., 2002), as the number of samples is imbalanced between the primary and metastasized group.

2.2 Deep learning framework

The first part of our model's framework comprises three key components, namely the AE (Hinton and Salakhutdinov, 2006), DeepLIFT (Shrikumar et al., 2017), and the deep neural network (DNN) (Svozil et al., 1997) (Figure 2).

First, the AE-based component is an unsupervised deep neural network with multiple stacked hidden layers composed of two parts, an encoder, and a decoder. The encoder maps the original (high-dimensional) data X to a reduced representation (100 dimensions) through the bottleneck layer. The purpose of the decoder is to reconstruct the original data $\hat{X}$ from the low-dimensional representation by minimizing the difference between X and $\hat{X}$. In this manner, the AE extracts features that differ from the original features and functions as a feature extraction method. We used the Python Keras library (https://github.com/fchollet/keras) to implement an AE consisting of three fully connected hidden

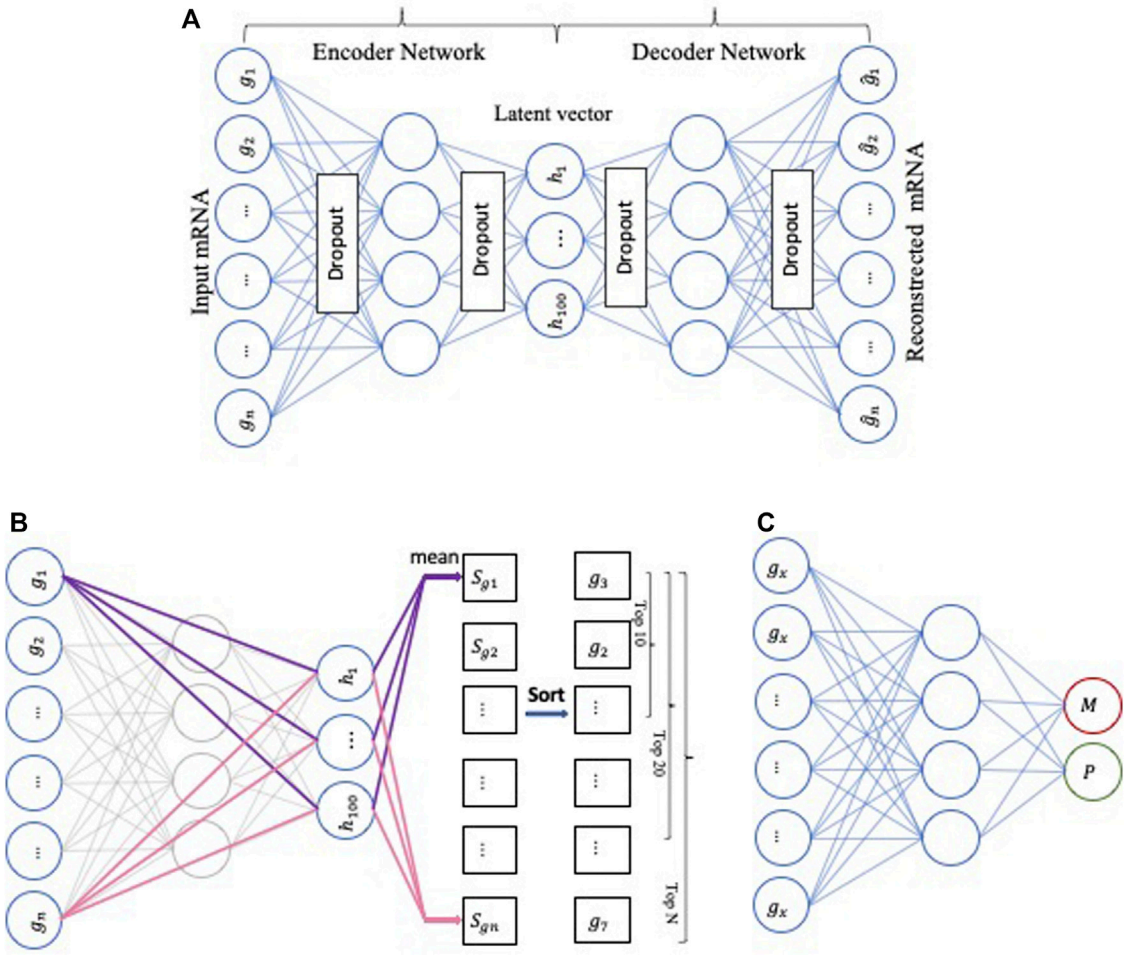


FIGURE 2
The workflow represents the first part of our model's framework. **(A)** The architecture of AutoEncoder, **(B)** Applying DeepLIFT to compute the importance scores in the Encoder network, **(C)** Using DNN as a baseline method to perform the metastasis prediction.

layers containing 500, 100, and 500 neurons. For each layer, we used "relu" as the activation function. Given m samples, each has a gene expression profile containing n genes; the input vector is reconstructed through a series of matrix transformations of multiple network layers. Training an AE involves finding parameters that minimize a specific loss function; we used mean absolute error (MAE) as the loss function. In addition, we added an L2 regularization penalty to control overfitting and used the early stopping technique. Finally, we trained the AE using the Adam (Kingma and Ba, 2014) optimization algorithm with 500 epochs and a 10% dropout.

Second, the DeepLIFT-based component is a feature scoring algorithm to calculate the contribution scores of each neuron. In our computing framework, we used DeepLIFT to calculate a contribution score for every gene of each input sample. The obtained contribution scores express the importance of the corresponding genes for the compression features of the low-dimensional representation (bottleneck) layer. Then, we ranked the genes based on their importance scores.

Third, the DNN-based component is a neural network with three hidden layers with 64, 32, and 8 neurons, respectively, and uses "relu" as the activation function. We used the Python Keras library to design the DNN model to predict if a sample is primary or metastasized. Finally, we iteratively added ten top-ranked genes (based on the importance scores) to train the DNN model.

The second part of our model uses the output from the first part, i.e., the most important genes for all sites, as an input to the final multi-class DNN model (Figure 3). This multi-class DNN consists of three hidden layers, each with 100 neurons, and uses a "relu" activation layer followed by an output layer with five output neurons (one for each class: primary, and metastasized to bone, brain, lung, or liver) that use the soft-max function to do the prediction. We then used the DeepLIFT to identify the most relevant neurons in each hidden layer for each of the five predictions (see 2.3 for details). Finally, the

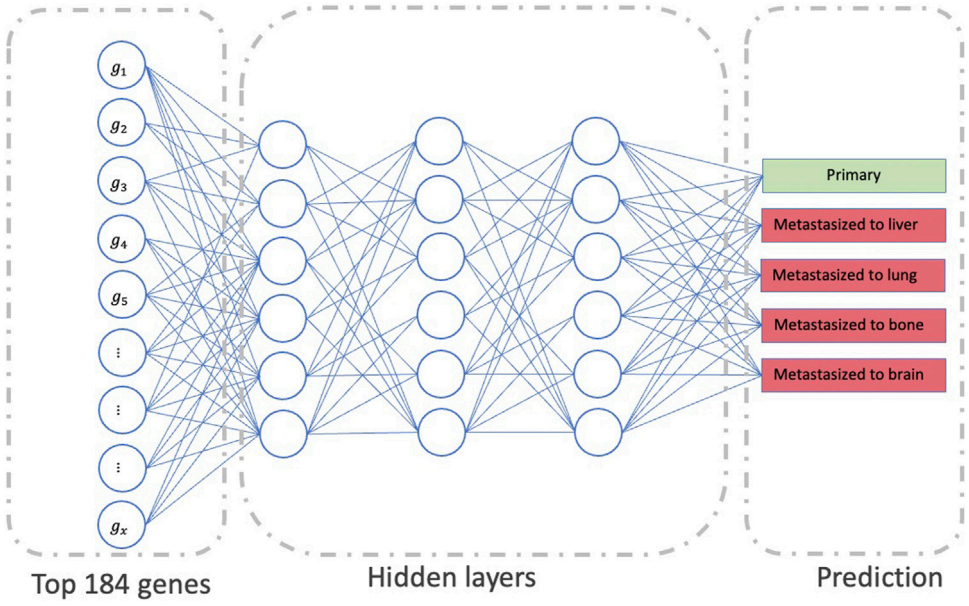


FIGURE 3
The workflow represents the second part of our model's framework, which determines the significant neurons in the network to predict metastasis status.

model was implemented in Python v.3.6 scripting language (https://www.python.org/), using the Keras deep learning and DeepLIFT frameworks (Figures 2B,C). Concerning time complexity, the time needed to train the model was 20.4 min for around 100 epochs for all samples using a workstation with Linux Ubuntu 18.04.5 LTS Intel Xeon Platinum 8,176, 64-bit OS and two GPUs: Quadro and Titan, with CUDA version 11.0.

2.3 Identifying the biological functions that the DL model uses to perform the prediction

We interpreted the prediction for each class by first computing the relevance scores through the DL network and identifying the most essential neurons that allow predicting the class. Then, we connected each important neuron with the list of the input genes affecting the neuron activation. In this manner, we associated biological functions with each layer based on its essential neurons.

The first step is to identify the neurons that most influence the predictions for each class (Bach et al., 2015; Hanczar et al., 2020). For this, we computed the relevance scores R of all neurons using the Deep-LIFT approach for each predicted class at each layer. Next, we used the mean of these relevance scores to obtain the average relevance of neuron i in layer L, representing this neuron's influence on the DL network to predict the class. The relevance score for neuron i in layer L is defined as the sum of incoming scores from each neuron j in layer $L+1$.

Finally, we ranked the neurons according to their average relevance scores and chose the most essential ones. Similar to (Hanczar et al., 2020), assuming that the average relevance scores follow a Gaussian distribution, we used the two-side t-test (p-value at 0.05) to determine each class's most essential neurons in each layer.

For a given important neuron in layer L, its activation is back propagated using the Deep-LIFT approach to compute the relevance score of each input gene. We then identified the most critical inputs that have an impact on the activation of the neuron. Similar to identifying the essential neurons, we used a two-sided t-test to select the essential input genes.

The second step is to connect each essential neuron to biological functions from GO, signature gene-set from MSigDB, and diseases from DisGeNET. Finally, we used an R interface to the Enrichr database EnricherR (Kuleshov et al., 2016) to identify the over-represented functions in the list of genes connected with each important neuron.

3 Results and discussion

3.1 Determining the gene set that provides optimal prediction performance

In the first part of our model's framework, we used GEO samples to train an AE and applied DeepLIFT to calculate importance scores for each gene for ranking. Then, ten top-ranked genes (based on the importance scores) were iteratively fed to the DNN model to identify

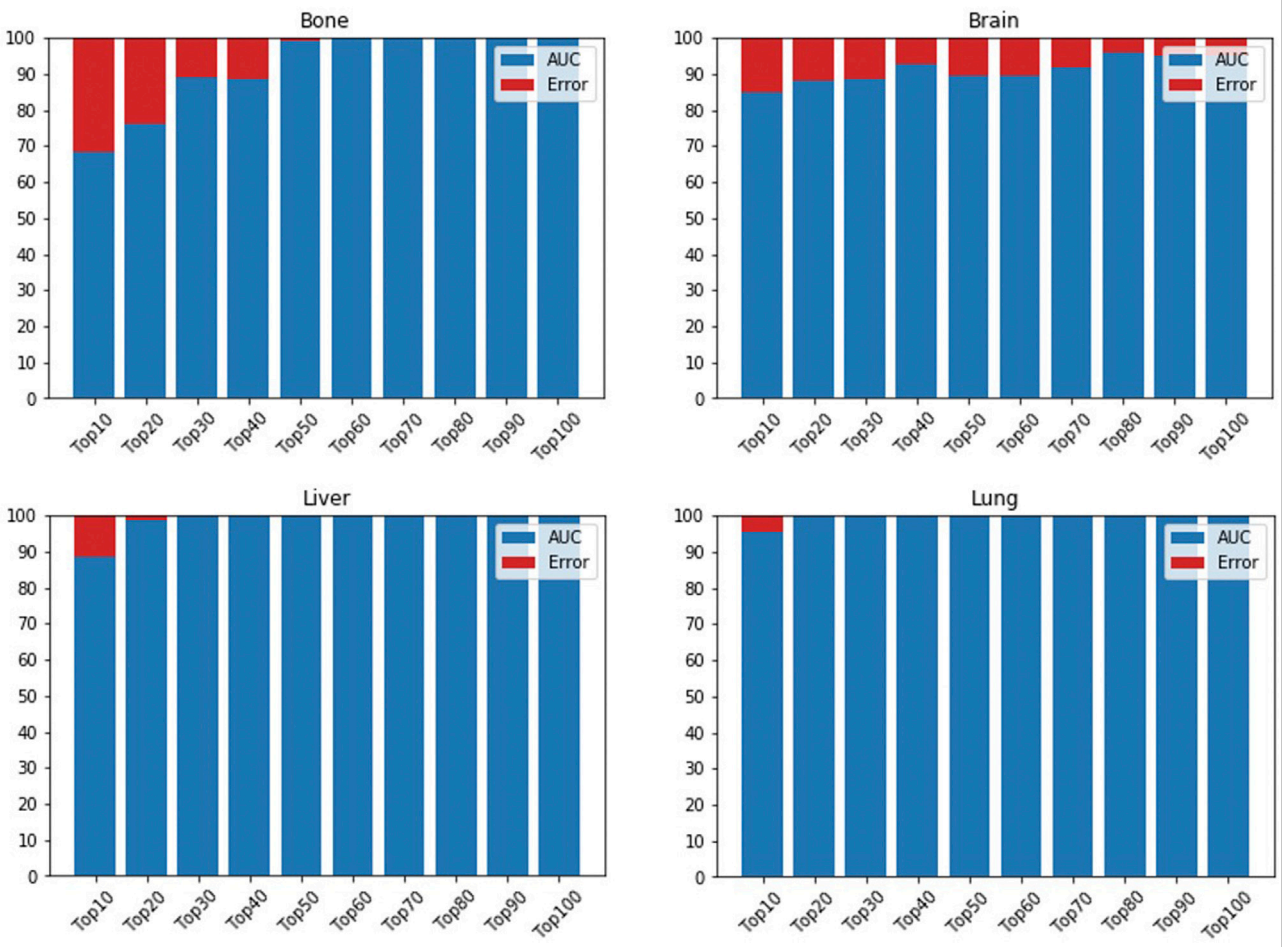


FIGURE 4
AUC is based on different numbers of featured genes using DNN for bone, brain, lung, and liver sites. AUC is indicated in blue, while error rate is shown in red.

the gene set that provides maximum performance when determining if a sample is primary or metastasized. The DNN reaches its maximum performance when including 60, 80, 20, and 30 top-ranked genes in metastasis to bone, brain, lung, and liver data, respectively (Figure 4). For the metastasis to bone, lung, and liver samples, the DNN achieved an AUC of 1.0. However, the DNN could only achieve an AUC of 0.9597 for the metastasis to brain samples. This might result from the brain samples having less than 50 primary samples, while the metastasis to bone, lung, and liver samples were analyzed using more than 200 primary samples. Adding weight to this suggestion is the number of metastasized samples used for each site being relatively the same (about 100 samples each).

3.2 Cross-site generalization analysis

After removing the duplicates, the 190 essential genes for all sites identified in the part of the model's framework were reduced to 184 genes. The 184 genes were used as an input to the final multi-class DNN model. This model takes these 184 genes and predicts if the input samples are primary or metastasized to the bone, brain, lung, or liver site. Figure 5 provides the prediction performance for the final multi-class DNN model. The best prediction performance was achieved for the primary samples (AUC of 0.93), followed by the metastasis to bone and lung samples (AUC of 0.88). The metastasis to liver and brain samples achieved lower prediction performances with an AUC of 0.84 and 0.82, respectively. Here, we expected the prediction performance for metastasis to the brain to be the lowest, based on the maximum performance the DNN achieved in Figure 3. However, the final multi-class DNN model achieved a more than acceptable prediction performance in all categories.

3.2.1 Testing the robustness of the final DNN model

The final multi-class DNN model achieved a good prediction performance; however, the prediction performance does not

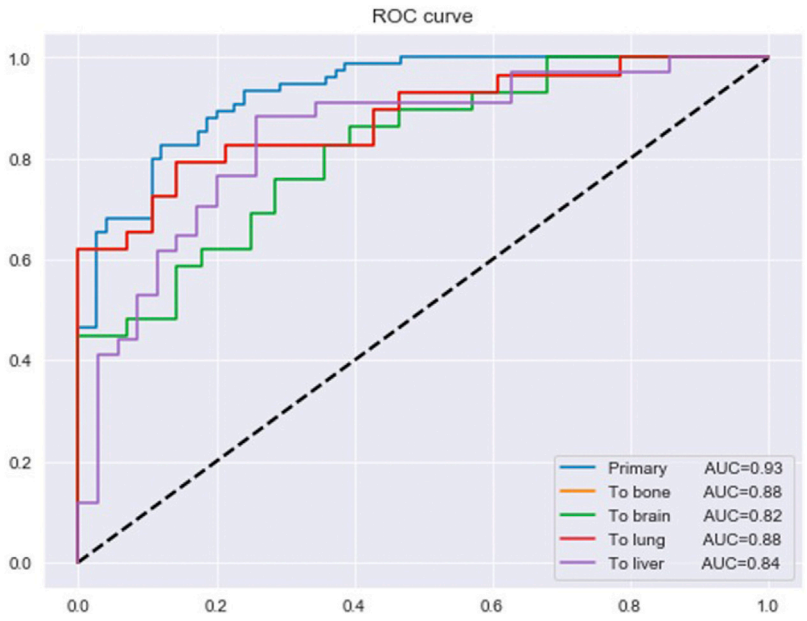
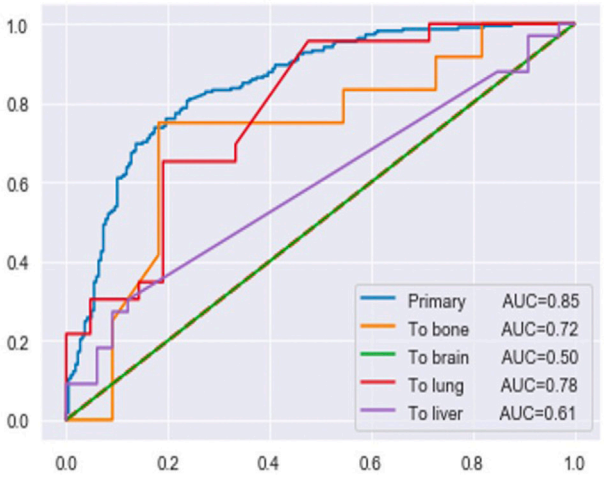


FIGURE 5
The prediction performance of the final multi-class DNN model.



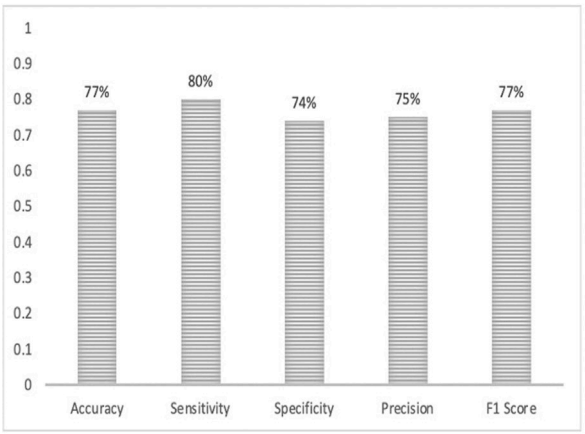


FIGURE 6
The prediction performance of the final multi-class DNN model using external testing data from the TCGA datasets. Note, for the brain there are only 2 samples in the test set).

indicate the robustness of the final DNN model. Thus, we further evaluated the model's performance using external testing data from the TCGA datasets (Figure 6) and using a population-based cohort (Figure 7). Also, by using external datasets as a validation technique to show how accurately our predictive model will perform in practice, we eliminate any concerns about over/under -fitting. First, the external set was extracted from the human cancer metastasis database (HCMDB) (Zheng et al., 2018), where we found 378 samples, 250 primary, and 21,2, 44, and 61 were

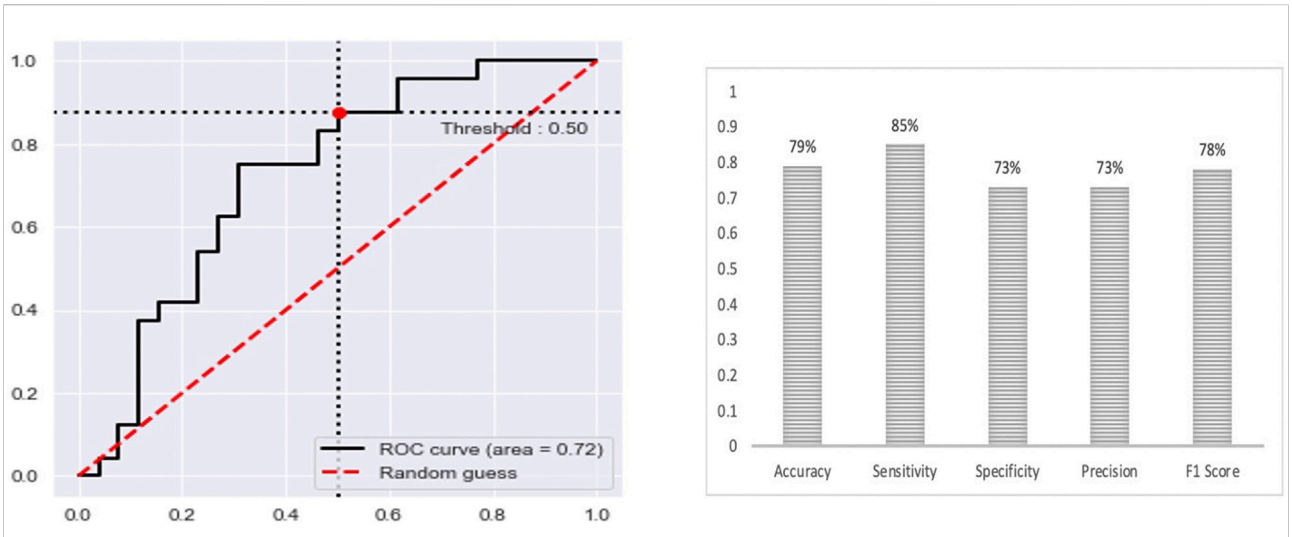


FIGURE 7
The prediction performance of the final multi-class DNN model using a specific population-based cohort.

metastasized to bone, brain, lung, and liver (see the complete list of TCGA IDs in Supplementary Table S1), respectively. In addition, we used the gene expression profiles of fresh breast cancer tissue of 45 (21 primary and 24 metastasized) Saudi-Arabian subjects deposited on GSE36295 to test the performance of our model on a population-based cohort (real data).

Figure 6 provides the prediction performance using the external set in terms of the area under the ROC curve and shows several other metrics, including accuracy, sensitivity, specificity, precision, and F1 score, ranging between 74%–80%. The prediction performance using the external data followed the same trend with the highest prediction performance achieved for the primary (AUC of 0.85) samples followed by the metastasis to lung (AUC of 0.78), bone (AUC of 0.72), liver (AUC of 0.61) and brain (AUC of 0.50) samples, respectively. This result shows that the multi-class model exhibits robustness concerning the three categories: the primary and metastasis to lung and bone samples. However, the prediction performance for the metastasis to the brain and liver samples dropped by 32% and 23 %, respectively. This suggests that we may have to re-establish the gene set that provides maximum performance using a larger cohort of samples (when the samples become available). Beyond that, here it should also be taken into consideration that for the brain we only had two samples in the test set.

Nonetheless, the prediction performance using samples from a population-based cohort shows that the multi-class DNN model achieved good prediction performance based on area under the ROC curve (AUC of 0.72), when distinguishing between the primary and metastatic samples (Figure 6), and shows several other metrics, including accuracy, sensitivity, specificity, precision, and F1 score, ranging between 73%–85%. This result gives an indication of the potential of our model to accurately predict metastasis sites.

3.2.2 The biological functions associated with the genes used by the DL model to perform the prediction

We further designed the model's workflow to provide a second functionality beyond metastasis site prediction, i.e., to identify the biological functions that the DL model uses to perform the prediction (Figure 8).

We achieved this through the biological interpretation of the neural network predicting the metastasis. That is, for each class, the essential neurons of each layer are selected based on the mean of the relevance scores by using the method described in Section 2.3. We identified 89, 56, 41, 16, and 53 essential neurons in hidden layer one for primary, metastasized to bone, brain, lung, and liver, respectively. We also identified 36, 40, 99, 11, and 18 essential neurons in hidden layer two for primary, metastasized to bone, brain, lung, and liver, respectively. Finally, we identified 54, 48, 22, 84, and 35 essential neurons in hidden layer three for primary, metastasized to bone, brain, lung, and liver, respectively.

For each essential neuron's list of genes, we determined GO biological functions based on a p-value < 0.05. Figure 9 provides the GO biological functions associated with the list of genes used to differentiate the primary samples from the metastasized ones. The critical neurons in each layer can be grouped depending on the functions enriched among the significant genes they contain. Overall, the enriched functions in layer one belonged to five main categories: "Metabolic process," "Cellular process," "Immune response," "Transport", and "Cell cycle." The enriched functions from the essential neurons of layer two included

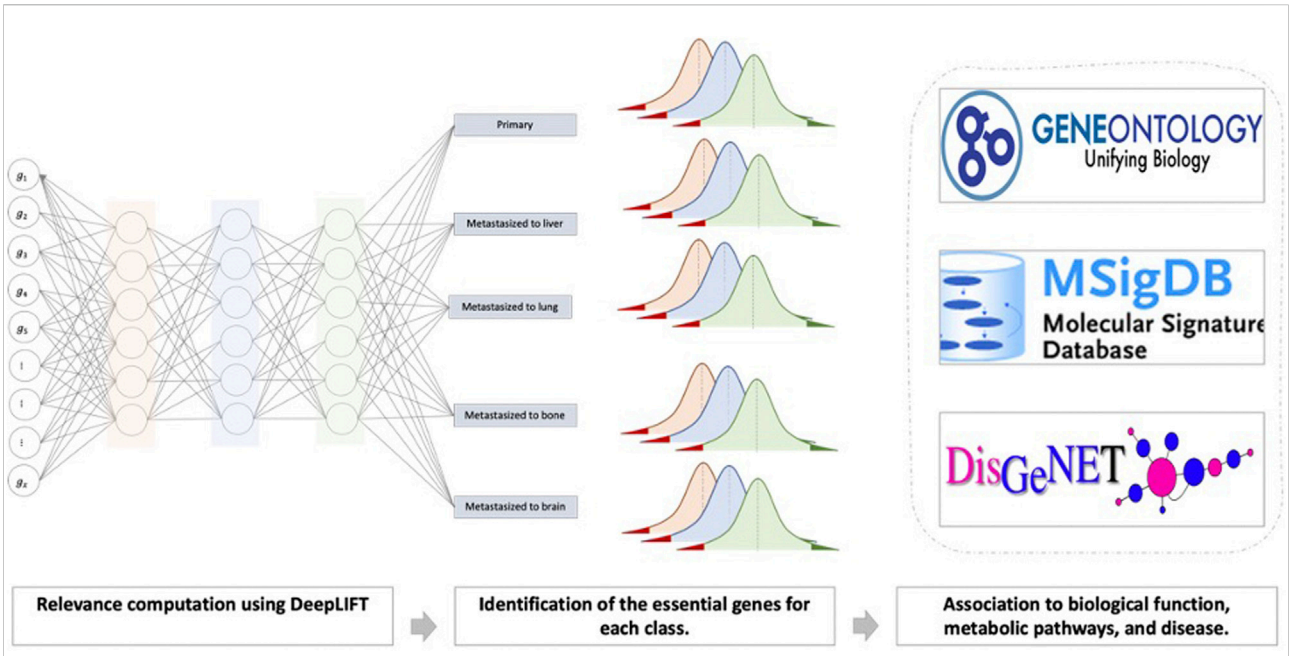


FIGURE 8
The biological interpretation of our deep neural network approach.

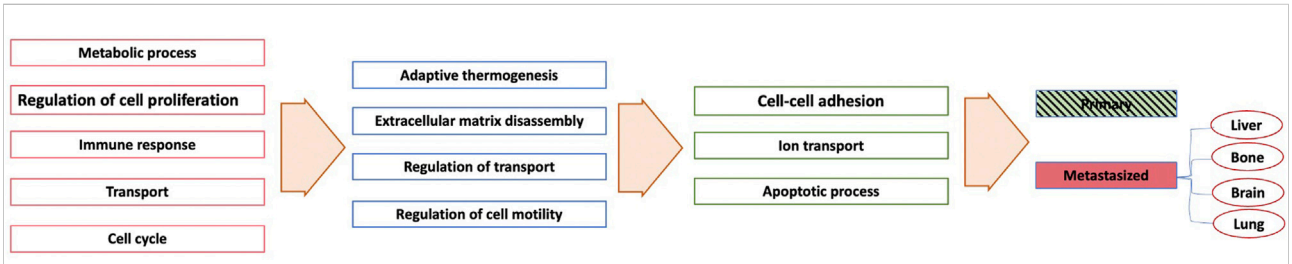


FIGURE 9
A simplified network showing each layer's enriched pathways based only on the metastasis sites.

"Adaptive thermogenesis," "Extracellular matrix disassembly," "Regulation of transport," and "Regulation of cell motility." The enriched functions from the essential neurons of layer three belonged to "Cell-cell adhesion," "Ion transport," "Apoptotic process." The first layer exhibits more general function categories but more specific functions are appearing in subsequent neural network layers.

To support and increase the biological insights extracted from the essential neurons' list of genes, we also performed MSigDB enrichment. In this analysis, we only considered MSigDB enrichments significant to at least three metastasis sites. Only three enriched categories were significant to all four sites, namely "Epithelial Mesenchymal Transition," "Apoptosis," and "IL-2/STAT5 Signaling" (Table 2). The apoptosis category is enriched based on the interpretation of the neural network (Figure 9) and the MSigDB enrichment. This is interesting as metastasis cells are subjected to various apoptotic stimuli and epithelial-mesenchymal transition (EMT) (which also features in the MSigDB enrichment) allows a polarized epithelial cell to undergo several biochemical changes to become a mesenchymal cell phenotype with enhanced resistance to apoptosis and increased migratory capacity and invasiveness and production of ECM components (Jason et al., 2003; Kalluri and Weinberg, 2009). Specifically, extracellular matrix disassembly (a GO biological function highlighted by the neural network) enzymes facilitates the remodeling of the extracellular matrix to create a microenvironment in the distant organ that promotes metastasis (Scheau et al., 2019; Winkler et al., 2020). Cell-cell adhesion, another GO

TABLE 2 MSigDB enrichment analysis.

	p-value			
MSigDB	**Bone**	**Brain**	**Lung**	**Liver**
Allograft rejection	2.60E-02	5.27E-03	Na	3.60E-03
Apoptosis	2.15E-02	2.00E-03	2.09E-02	2.15E-02
Coagulation	1.41E-02	na	1.82E-02	1.88E-02
DNA repair	1.53E-02	na	1.27E-02	2.02E-02
Epithelial mesenchymal transition	2.60E-02	1.98E-02	1.66E-02	2.60E-02
Glycolysis	2.60E-02	na	2.53E-02	3.11E-02
IL-2/STAT5 signaling	2.59E-02	2.02E-03	2.52E-02	2.59E-02
Interferon alpha response	na	1.39E-03	8.68E-04	1.39E-03
mTORC1 signaling	2.60E-02	na	1.66E-02	2.60E-02
Oxidative phosphorylation	3.60E-03	na	1.37E-03	3.28E-04
UV response up	2.12E-02	1.63E-02	2.06E-02	Na

[a]Enrichment associations that are significant to at least three sites

biological function highlighted by the neural network, is also a key element of metastasis. For example, it has been shown that S100A8/A9 from tumor cells bind to RAGE on myeloid-derived suppressor cells (MDSCs) and promotes the migration and accumulation of MDSC, while periostin from MDSCs participates in pre-metastatic niche (PMN) formation through promoting extracellular matrix remodeling to facilitate the metastatic colonization of disseminated tumor cells (Cheng et al., 2008; Sinha et al., 2008; Wang et al., 2016). Overall, these biological functions suggest the gene lists used in the DL model to perform the prediction are to a large extent metastasis-specific and can be used to retrieve metastasis-specific biological functions beyond its metastasis site prediction capabilities (Sinha et al., 2008).

We also performed DisGenNET enrichment. In this analysis, we only considered DisGenNET disease enrichments significant to at least two metastasis sites. Only eight enriched disease categories were significantly associated with at least two metastasis sites, namely Autoimmune Diseases, Carcinoma breast stage IV, Cirrhosis, Dermatomyositis, Giant Cell Tumors, Leukemia, Metastatic malignant neoplasm to brain, and Rheumatoid Arthritis (Table 3). Four disease categories are associated with cancer, and noteworthy is the late-stage and metastasized cancer that is being picked up. Beyond this, Dermatomyositis (Luu et al., 2015), Rheumatoid Arthritis (Racanelli et al., 2008), and Autoimmune Diseases (Milkiewicz et al., 1999) are recognized paraneoplastic syndromes, which are symptoms that occur at sites distant from a tumor or its metastasis site (Pelosof and Gerber, 2010). In addition, several of our differentially expressed genes, including HLA-DMA, SOCS1, HLA-C, CTNNB1, KRAS, MET, and CD244, are associated with Liver Cirrhosis (Knouse et al., 2019), CD79A, HLA-DMA, SOCS1, HLA-B, HLA-C, IFI35, CD68, MET, PTHLH, CD244, and C2 with Rheumatoid Arthritis (Roy et al., 2011), HLA-B, HPRT1, and C2 with dermatomyositis (Bonnetblanc et al., 1990), and RB1, HLA-DMA, CXCR4, and CTGF with Cirrhosis (Shah and Casciola-Rosen, 2015). We also have several genes, including FTO, HLA-DMA, GAP43, SCN8A, HLA-C, CD68, and CDR2, associated with Multiple Sclerosis (Plantone et al., 2015), which suggest Multiple Sclerosis and Cirrhosis may possibly be a paraneoplastic syndrome that arises with metastasis.

We further determined the overlapping genes between the primary and metastasis samples for the four sites. This analysis includes only the genes used by the DL to perform the classification. If we only considered genes common to at least three sites, we found the products of two genes, HIP1 and LARP4, with expression levels downregulated in the primary samples but upregulated in the metastasis samples. HIP1 was used by the DL to predict metastasis to the bone, brain, and lung, while LARP4 was used to predict metastasis to the brain, lung, and liver. This is interesting as HIP1 is one of the essential proteins involved in clathrin-mediated endocytosis (CME) (Chang et al., 2015), and crosstalk between CLCb/Dyn1-mediated adaptive CME and epidermal growth factor receptor (EGFR) signaling increases metastasis (Chen et al., 2017). Also, LARP4, a known RNA-binding protein (RBP) (Yang et al., 2011; Chothani et al., 2019) that repress or activate the translation of target genes, change the cell shape (which has been correlated with metastatic potential) and LARP4 depletion increases cell migration and invasion (Lyons et al., 2016; Seetharaman et al., 2016). Other proteins also upregulated and common to at least three sites (but do not appear in the primary samples gene list) include CC2D1A (Kumar et al., 2019), CD68 (Huang et al., 2018), EFCAB1 (Fagone et al., 2017), HLA-DMA (Li et al., 2020a), PRAME (Huang et al., 2016; Al-Khadairi et al., 2019), and ULBP2 (Paschen et al., 2009), all of which was linked to

TABLE 3 DisGeNET enrichment analysis.

	p-value			
DisGeNET	**Bone**	**Brain**	**Lung**	**Liver**
Autoimmune Diseases	5.12E-04	na	6.43E-05	Na
Carcinoma breast stage IV	2.68E-05	na	4.90E-04	Na
Cirrhosis	3.80E-04	na	0.00E+00	1.37E-02
Dermatomyositis	2.57E-05	2.57E-05	Na	Na
Giant Cell Tumors	5.69E-06	na	7.84E-08	Na
leukemia	7.13E-05	7.13E-05	Na	Na
Metastatic malignant neoplasm to brain	1.84E-04	na	3.43E-07	Na
Rheumatoid Arthritis	4.84E-05	na	2.71E-06	Na

[a]Disease associations that are significant to at least two sites

metastasis in previous studies. In fact, 87 % of the essential genes are associated with metastasis-related functions based on the current literature (Table 4).

4 Concluding remarks

Metastasis remains the leading cause of cancer-related deaths worldwide, and our inability to identify the tumor cells colonizing distant sites means that the physician cannot treat the metastasized tumors. Here, we developed a DL model that can be fed raw gene expression data to predict whether a sample is primary or metastasized to the brain, bone, lung, or liver. The final multi-class DNN model achieved more than acceptable prediction performance in all categories. We achieved the best prediction performance for the primary samples (AUC of 0.93), followed by the metastasis to bone and lung samples (AUC of 0.88). On the other hand, the metastasis to liver and brain samples achieved lower prediction performance with an AUC of 0.84 and 0.82, respectively. We observed the same trend when evaluating the prediction performance using external data, i.e., the highest prediction performance for the primary (AUC of 0.85) samples followed by the metastasis to lung (AUC of 0.78), bone (AUC of 0.72), liver (AUC of 0.61) and brain (AUC of 0.50) samples, respectively. However, the prediction performance for the metastasis to the brain and liver samples dropped by 32% and 23 %, respectively.

Many factors may contribute to the result we obtained for the brain samples, as this data had the highest number of DEGs and required the highest amount of top-ranked genes to be included in the model, indicating biological complexity associated with the metastasis to the brain. Additionally, the brain samples had less than 50 primary samples. In contrast, we analyzed the metastasis to bone, lung, and liver samples using more than 200 primary samples (the number of metastasized samples used for each site was similar, about 100 samples each). Beyond that, the brain only had two samples in the test set for the external data that exhibited the massive drop in prediction performance. Having this lower number of brain samples may also be contributing to the much lower prediction performance achieved with it. Thus, in the future, we will re-establish the gene set that provides maximum performance using a larger cohort of samples (when the data become available). Nonetheless, we further evaluated the prediction performance using samples from a population-based cohort to show that the multi-class DNN model achieved good prediction performance (AUC of 0.72) when distinguishing between the primary and metastatic samples, which shows the potential of our model.

We further designed the model's workflow to provide a second functionality beyond metastasis site prediction, i.e., to identify the biological functions that the DL model uses to perform the prediction. We achieved this by associating GO biological functions (p-value < 0.05) with the neuron's list of genes that differentiate the primary samples from the metastasized ones in the DL model. The critical neurons in each layer are grouped depending on the functions enriched. Thus, the first layer exhibits more general function categories, but more specific functions appear in subsequent neural network layers. Finally, we compared the enrichments retrieved through the DL model neuron interpretations with the MSigDB enrichment analysis. We found only a few functional categories common to both analyses but several inter-related categories. For example, the literature shows "Epithelial Mesenchymal Transition' involves 'Ion transport," and "Extracellular matrix disassembly," and it is linked to "Cell-cell adhesion," "regulation of cell motility" and "apoptosis process" (Jason et al., 2003; Cheng et al., 2008; Sinha et al., 2008; Kalluri and Weinberg, 2009; Wang et al., 2016; Scheau et al., 2019; Winkler et al., 2020). Overall, these biological functions suggest that the gene lists used in the DL model to perform the prediction are to a large extent metastasis-specific, which is further supported by literature showing 87% of the genes used by the DL have already been linked to metastasis. These results clearly suggest that our DL model can be used to retrieve metastasis-specific biological functions beyond its metastasis site prediction capabilities.

TABLE 4 Literature linking the genes used by the DL to metastasis.

Gene	Link	Gene	Link	Gene	Link	Gene	Link
ACTC1	Ohtaki et al. (2017)	FTO	Ding et al. (2020)	NDUFC2-KCTD14	NA	RUBCN	Marsh and Debnath, (2020)
ADAM10	Xu et al. (2010)	GABARAP	Liu et al. (2021a)	NF1	Kitamura et al. (2010)	SCIN	NA
ANO1	Zhang et al. (2021a)	GAP43	Zhang et al. (2018a)	NOL3	Medina-Ramirez et al. (2011)	SCLY	Hartung et al. (2017)
ATP5PD	Song et al. (2016)	GAPDHS	Liu et al. (2017)	OCLN	Wang et al. (2018a)	SCN8A	(Hartung et al. (2017); Lopez-Charcas et al. (2018))
ATP5PO	McLaren and University of Western Australia, (2009)	GINS3	Li et al. (2021)	PACS2	Madreiter-Sokolowski et al. (2021)	SIGLEC1	Strömvall et al. (2017)
C2	NA	GNL3L	Kannathasan et al. (2020)	PCNX2	Yamaguchi et al. (2016)	SLC6A16	Nałęcz, (2020)
C5orf22	Schulten et al. (2017)	HIP1	Sun et al. (2021)	PFAS	Lv et al. (2020)	SNORD107	Xu et al. (2016)
C7orf25	NA	HIP2	Wu et al. (2020)	PIAS1	Wang et al. (2018b)	SNORD19B	Xu et al. (2016)
CC2D1A	Kumar et al. (2019)	HIP3	NA	PRAME	Huang et al. (2016)	SNORD42A	NA
CD244	Johnson et al. (2003)	HLA-B	(Cordon-Cardo et al. (1991); Jiang et al. (2014))	PRKACA	Honeyman et al. (2014)	SOCS1	David et al. (2014)
CD68	Huang et al. (2018)	HLA-C	Cordon-Cardo et al. (1991)	PRR14	Li et al. (2019a)	SSH3	Hu et al. (2019a)
CD79A	Luger et al. (2013)	HLA-DMA	Li et al. (2020a)	PTHLH	(Li et al. (2019a); Pitarresi et al. (2021))	SSX1	NA
CD82	Di Giacomo et al. (2017)	HPRT1	J Sedano et al. (2020)	RAB15	Iacobas et al. (2018)	ST20-MTHFS	NA
CD83	Giorello et al. (2021)	HPS4	Liu et al. (2018)	RAB26	Liu et al. (2021b)	SUOX	Yano et al. (2021)
CDR1	Harrison et al. (2020)	HSPA9	Yi et al. (2008)	RAD51B	Seguin et al. (2018)	TEP1	(Hwang et al. (2001); Yano et al. (2021))
CDR2	Balamurugan et al. (2009)	IFI35	Hu et al. (2021)	RB1	Ku et al. (2017)	TMSB4Y	Wong et al. (2015)
CHD1L	He et al. (2012)	IFITM2	Xu et al. (2017)	RHOB	Ju et al. (2020)	TPT1P8	NA
CHRNA1	Chang et al. (2013)	JAM3	(Xu et al. (2017); Zhou et al. (2019a))	RHOBTB2	Ling et al. (2010)	TREX1	Feng et al. (2016)
CTGF	Okusha et al. (2020)	KRAS	Boutin et al. (2017)	RPL13	Ebright et al. (2020)	TUBA3C	Zhou et al. (2019b)
CTNNB1	Wen et al. (2019)	KRT1	Han et al. (2021)	RPL21	Li et al. (2020b)	TUBGCP3	NA
CXCR4	Zhang et al. (2021b)	LARP4	Egiz et al. (2019)	RPL9	Baik et al. (2016)	UBA6	Cheng et al. (2021)
EFCAB1	Fagone et al. (2017)	LCP1	Ge et al. (2020)	RPP30	NA	UBD	Cheng et al. (2021)
ERCC3	Zhang et al. (2020)	LDHAL6B	(Ge et al., 2020; Liu et al., 2020)	RPS24	Wang et al. (2020)	ULBP2	Cheon et al. (2011)
ESR2	Song et al. (2018)	MET	Zhang et al. (2018b)	RPS6KA2	NA	USP6	Zeng et al. (2018)
FAM153A	NA	MME	Li et al. (2019b)	RPS8	Mao-De and Jing, (2007)	ZNF236	NA
FAXDC2	NA	MMP	Gonzalez-Avila et al. (2019)	RRP12	Hu et al. (2019b)	ZNF764	NA
FGF23	Ewendt et al. (2020)	NBEAL2	Rae et al. (2015)				

5 Availability

We also developed a web server that the scientific community can access. The web-based tool, MetastaSite https://www.cbrc.kaust.edu.sa/metastasite/, provides a means to implement the final multi-class DNN model developed in the current study. It allows the users to predict the metastasis site (primary, metastasized to bone, brain, lung, or liver). The user needs to provide the raw gene expression for every sample.

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

SA, ME, and XG conceived and designed the study; SA performed the experiments; SA, MT, AsA, AbA, SA, and ME analyzed the results; MU designed the web tool; SA, AsA, AbA, SA, MU, MT, TG, ME, and XG contributed to writing and reviewing the manuscript. All authors read and approved the final manuscript.

Funding

The research reported in this publication was supported by King Abdullah University of Science and Technology (KAUST) through grant awards Nos. BAS/1/1059-01-01, BAS/1/1624-01-01, FCC/1/1976-20-01, FCC/1/1976-26-01, URF/1/3450-01-01, REI/1/4216-01-01, REI/1/4437-01-01, REI/1/4473-01-01, and URF/1/4098-01-01.

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

Al-Khadairi, G., Naik, A., Thomas, R., Al-Sulaiti, B., Rizly, S., and Decock, J. (2019). PRAME promotes epithelial-to-mesenchymal transition in triple negative breast cancer. *J. Transl. Med.* 17 (1), 9. doi:10.1186/s12967-018-1757-3

Albaradei, S., Napolitano, F., Thafar, M. A., Gojobori, T., Essack, M., and Gao, X. (2021). MetaCancer: A deep learning-based pan-cancer metastasis prediction model developed using multi-omics data. *Comput. Struct. Biotechnol. J.* 19, 4404–4411. doi:10.1016/j.csbj.2021.08.006

Albaradei, S., Thafar, M. A., Van Neste, C., and Essack, M. (2019). "Metastatic state of colorectal cancer can be accurately predicted with methylome," in Proceedings of the 2019 6th International Conference on Bioinformatics Research and Applications, Seoul Republic of Korea, December 19 - 21, 2019.

Albaradei, S., Thafar, M., Alsaedi, A., Van Neste, C., Gojobori, T., Essack, M., et al. (2021). Machine learning and deep learning methods that use omics data for metastasis prediction. *Comput. Struct. Biotechnol. J.* 19, 5008–5018. doi:10.1016/j.csbj.2021.09.001

Albaradei, S., Uludag, M., Thafar, M. A., Gojobori, T., Essack, M., and Gao, X. (2021). Predicting bone metastasis using gene expression-based machine learning models. *Front. Genet.* 12, 771092. doi:10.3389/fgene.2021.771092

Bach, S., Binder, A., Montavon, G., Klauschen, F., Muller, K. R., and Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* 10 (7), e0130140. doi:10.1371/journal.pone.0130140

Baik, I. H., Jo, G. H., Seo, D., Ko, M. J., Cho, C. H., Lee, M. G., et al. (2016). Knockdown of RPL9 expression inhibits colorectal carcinoma growth via the inactivation of Id-1/NF-κB signaling axis. *Int. J. Oncol.* 49 (5), 1953–1962. doi:10.3892/ijo.2016.3688

Balamurugan, K., Luu, V. D., Kaufmann, M. R., Hofmann, V. S., Boysen, G., Barth, S., et al. (2009). Onconeuronal cerebellar degeneration-related antigen, Cdr2, is strongly expressed in papillary renal cell carcinoma and leads to attenuated hypoxic response. *Oncogene* 28 (37), 3274–3285. doi:10.1038/onc.2009.186

Bonnetblanc, J. M., Bernard, P., and Fayol, J. (1990). Dermatomyositis and malignancy. *Dermatology* 180 (4), 212–216. doi:10.1159/000248032

Boutin, A. T., Liao, W. T., Wang, M., Hwang, S. S., Karpinets, T. V., Cheung, H., et al. (2017). Oncogenic Kras drives invasion and maintains metastases in colorectal cancer. *Genes Dev.* 31 (4), 370–382. doi:10.1101/gad.293449.116

Chang, P. M.-H., Yeh, Y. C., Chen, T. C., Wu, Y. C., Lu, P. J., Cheng, H. C., et al. (2013). High expression of CHRNA1 is associated with reduced survival in early stage lung adenocarcinoma after complete resection. *Ann. Surg. Oncol.* 20 (11), 3648–3654. doi:10.1245/s10434-013-3034-2

Chang, Y.-H., Chen, C. M., Chen, H. Y., and Yang, P. C. (2015). Pathway-based gene signatures predicting clinical outcome of lung adenocarcinoma. *Sci. Rep.* 5, 10979. doi:10.1038/srep10979

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi:10.1613/jair.953

Chen, P.-H., Bendris, N., Hsiao, Y. J., Reis, C. R., Mettlen, M., Chen, H. Y., et al. (2017). Crosstalk between CLCb/dyn1-mediated adaptive clathrin-mediated endocytosis and epidermal growth factor receptor signaling increases metastasis. *Dev. Cell* 40 (3), 278–288. doi:10.1016/j.devcel.2017.01.007

Cheng, F., Liu, J., Zhang, Y., You, Q., Chen, B., Cheng, J., et al. (2021). Long non-coding RNA UBA6-AS1 promotes the malignant properties of glioblastoma by competitively binding to microRNA-760 and enhancing homeobox A2 expression. *Cancer Manag. Res.* 13, 379–392. doi:10.2147/CMAR.S287676

Cheng, P., Corzo, C. A., Luetteke, N., Yu, B., Nagaraj, S., Bui, M. M., et al. (2008). Inhibition of dendritic cell differentiation and accumulation of myeloid-derived suppressor cells in cancer is regulated by S100A9 protein. *J. Exp. Med.* 205 (10), 2235–2249. doi:10.1084/jem.20080132

Cheon, S., Lee, J. H., Park, S., Bang, S. I., Lee, W. J., Yoon, D. Y., et al. (2011). Overexpression of IL-32alpha increases natural killer cell-mediated killing through up-regulation of Fas and UL16-binding protein 2 (ULBP2) expression in human chronic myeloid leukemia cells. *J. Biol. Chem.* 286 (14), 12049–12055. doi:10.1074/jbc.M110.159756

Chereda, H., Bleckmann, A., Kramer, F., Leha, A., and Beissbarth, T. (2019). Utilizing molecular network information via Graph convolutional neural networks to predict metastatic event in breast cancer. *Stud. Health Technol. Inf.* 267, 181–186. doi:10.3233/SHTI190824

Chothani, S., Schafer, S., Adami, E., Viswanathan, S., Widjaja, A. A., Langley, S. R., et al. (2019). Widespread translational control of fibrosis in the human heart by RNA-binding proteins. *Circulation* 140 (11), 937–951. doi:10.1161/CIRCULATIONAHA.119.039596

Cordon-Cardo, C., Fuks, Z., DrobnjakM.Moreno, C., Eisenbach, L., and FeldManM. (1991). Expression of HLA-A, B, C antigens on primary and metastatic tumor cell populations of human carcinomas. *Cancer Res.* 51 (23 Pt 1), 6372–6380.

Daoud, M., and Mayo, M. (2019). A survey of neural network-based cancer prediction models from microarray data. *Artif. Intell. Med.* 97, 204–214. doi:10.1016/j.artmed.2019.01.006

David, M., Naudin, C., Letourneur, M., Polrot, M., Renoir, J. M., Lazar, V., et al. (2014). Suppressor of cytokine signaling 1 modulates invasion and metastatic

potential of colorectal cancer cells. *Mol. Oncol.* 8 (5), 942–955. doi:10.1016/j.molonc.2014.03.014

Di Giacomo, V., Tian, T. V., MAs, A., PecoraroM.BatLLe-Morera, L., Noya, L., et al. (2017). ΔNp63α promotes adhesion of metastatic prostate cancer cells to the bone through regulation of CD82. *Oncogene* 36 (31), 4381–4392. doi:10.1038/onc.2017.42

Ding, Y., Qi, N., Wang, K., Huang, Y., Liao, J., Wang, H., et al. (2020). FTO facilitates lung adenocarcinoma cell progression by activating cell migration through mRNA demethylation. *Onco. Targets. Ther.* 13, 1461–1470. doi:10.2147/OTT.S231914

Ebright, R. Y., Lee, S., Wittner, B. S., Niederhoffer, K. L., Nicholson, B. T., Bardia, A., et al. (2020). Deregulation of ribosomal protein expression and translation promotes breast cancer metastasis. *Science* 367 (6485), 1468–1473. doi:10.1126/science.aay0939

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30 (1), 207–210. doi:10.1093/nar/30.1.207

Egiz, M., Usui, T., Ishibashi, M., Zhang, X., Shigeta, S., Toyoshima, M., et al. (2019). La-related protein 4 as a suppressor for motility of ovarian cancer cells. *Tohoku J. Exp. Med.* 247 (1), 59–67. doi:10.1620/tjem.247.59

Ewendt, F., Feger, M., and Föller, M. (2020). Role of fibroblast growth factor 23 (FGF23) and αKlotho in cancer. *Front. Cell Dev. Biol.* 8, 601006. doi:10.3389/fcell.2020.601006

Fagone, P., Caltabiano, R., Russo, A., Lupo, G., Anfuso, C. D., Basile, M. S., et al. (2017). Identification of novel chemotherapeutic strategies for metastatic uveal melanoma. *Sci. Rep.* 7, 44564. doi:10.1038/srep44564

Fakoor, R., Ladhak, F., Nazi, A., and Huber, M. (2013). "Using deep learning to enhance cancer diagnosis and classification," in Proceedings of the international conference on machine learning, New York, USA, 18-24 July 2021.

Feng, J., Lan, R., Cai, G., Lin, J., Wang, X., Lin, J., et al. (2016). Verification of TREX1 as a promising indicator of judging the prognosis of osteosarcoma. *J. Orthop. Surg. Res.* 11 (1), 150. doi:10.1186/s13018-016-0487-6

Ge, X., Liu, W., Zhao, W., Feng, S., Duan, A., Ji, C., et al. (2020). Exosomal transfer of LCP1 promotes osteosarcoma cell tumorigenesis and metastasis by activating the JAK2/STAT3 signaling pathway. *Mol. Ther. Nucleic Acids* 21, 900–915. doi:10.1016/j.omtn.2020.07.025

Giorello, M. B., Matas, A., Marenco, P., Davies, K. M., Borzone, F. R., Calcagno, M. d. L., et al. (2021). CD1a- and CD83-positive dendritic cells as prognostic markers of metastasis development in early breast cancer patients. *Breast Cancer* 28 (6), 1328–1339. doi:10.1007/s12282-021-01270-9

Gonzalez-Avila, G., Sommer, B., Mendoza-Posada, D. A., Ramos, C., Garcia-Hernandez, A. A., and Falfan-Valencia, R. (2019). Corrigendum to "Matrix metalloproteinases participation in the metastatic process and their diagnostic and therapeutic applications in cancer" [Crit. Rev. Oncol. Hematol. 137, May (2019) 57-83]. *Crit. Rev. Oncol. Hematol.* 138, 172. doi:10.1016/j.critrevonc.2019.04.017

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. Cambridge, Massachusetts, United States: MIT Press.

Grapov, D., Fahrmann, J., Wanichthanarak, K., and Khoomrung, S. (2018). Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. *OMICS A J. Integr. Biol.* 22 (10), 630–636. doi:10.1089/omi.2018.0097

Han, W., Hu, C., Fan, Z. J., and Shen, G. L. (2021). Transcript levels of keratin 1/5/6/14/15/16/17 as potential prognostic indicators in melanoma patients. *Sci. Rep.* 11 (1), 1023. doi:10.1038/s41598-020-80336-8

Hanczar, B., Zehraoui, F., Issa, T., and Arles, M. (2020). Biological interpretation of deep neural network for phenotype prediction based on gene expression. *BMC Bioinforma.* 21 (1), 501. doi:10.1186/s12859-020-03836-4

Harrison, E. B., Porrello, A., Bowman, B. M., Belanger, A. R., Yacovone, G., Azam, S. H., et al. (2020). A circle RNA regulatory Axis promotes lung squamous metastasis via CDR1-mediated regulation of golgi trafficking. *Cancer Res.* 80 (22), 4972–4985. doi:10.1158/0008-5472.CAN-20-1162

Hartung, F., Wang, Y., Aronow, B., and Weber, G. F. (2017). A core program of gene expression characterizes cancer metastases. *Oncotarget* 8 (60), 102161–102175. doi:10.18632/oncotarget.22240

He, W.-P., Zhou, J., Cai, M. Y., Xiao, X. S., Liao, Y. J., Kung, H. F., et al. (2012). CHD1L protein is overexpressed in human ovarian carcinomas and is a novel predictive biomarker for patients survival. *BMC Cancer* 12, 437. doi:10.1186/1471-2407-12-437

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313 (5786), 504–507. doi:10.1126/science.1127647

Honeyman, J. N., Simon, E. P., Robine, N., Chiaroni-Clarke, R., Darcy, D. G., Lim, I. I. P., et al. (2014). Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science* 343 (6174), 1010–1014. doi:10.1126/science.1249484

Hu, M., Fu, X., Si, Z., Li, C., Sun, J., Du, X., et al. (2019). Identification of differently expressed genes associated with prognosis and growth in colon adenocarcinoma based on integrated bioinformatics analysis. *Front. Genet.* 10, 1245. doi:10.3389/fgene.2019.01245

Hu, Y.-H., Lu, Y. X., Zhang, Z. Y., Zhang, J. M., Zhang, W. J., Zheng, L., et al. (2019). SSH3 facilitates colorectal cancer cell invasion and metastasis by affecting signaling cascades involving LIMK1/Rac1. *Am. J. Cancer Res.* 9 (5), 1061–1073.

Hu, Y., Wang, B., Yi, K., Lei, Q., Wang, G., and Xu, X. (2021). IFI35 is involved in the regulation of the radiosensitivity of colorectal cancer cells. *Cancer Cell Int.* 21 (1), 290. doi:10.1186/s12935-021-01997-7

Huang, Q., Wei, H., Wu, Z., Li, L., Yao, L., Sun, Z., et al. (2016). Preferentially expressed antigen of melanoma prevents lung cancer metastasis. *PLoS One* 11 (7), e0149640. doi:10.1371/journal.pone.0149640

Huang, Y., Pan, L., Helou, K., Xia, Q., Parris, T. Z., Li, H., et al. (2018). Mechanical ventilation promotes lung metastasis in experimental 4T1 breast cancer lung-metastasized models. *Cancer Manag. Res.* 10, 545–555. doi:10.2147/CMAR.S142650

Hwang, P. H., Yi, H. K., Kim, D. S., Nam, S. Y., Kim, J. S., and Lee, D. Y. (2001). Suppression of tumorigenicity and metastasis in B16F10 cells by PTEN/MMAC1/TEP1 gene. *Cancer Lett.* 172 (1), 83–91. doi:10.1016/s0304-3835(01)00632-2

Iacobas, D. A., Tuli, N. Y., Iacobas, S., Rasamny, J. K., Moscatello, A., Geliebter, J., et al. (2018). Gene master regulators of papillary and anaplastic thyroid cancers. *Oncotarget* 9 (2), 2410–2424. doi:10.18632/oncotarget.23417

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4 (2), 249–264. doi:10.1093/biostatistics/4.2.249

J Sedano, M., I Ramos, E., Choudhari, R., L Harrison, A., Subramani, R., Lakshmanaswamy, R., et al. (2020). Hypoxanthine phosphoribosyl transferase 1 is upregulated, predicts clinical outcome and controls gene expression in breast cancer. *Cancers* 12 (6), E1522. doi:10.3390/cancers12061522

Jason, L. T., George, N. N., and Ann, F. C. (2003). The role of apoptosis in tumor progression and metastasis. *Curr. Mol. Med.* 3 (7), 631–642. doi:10.2174/1566524033479483

Jiang, Y.-Z., Li, Q. H., Zhao, J. Q., and Lv, J. J. (2014). Identification of a novel fusion gene (HLA-E and HLA-B) by RNA-seq analysis in esophageal squamous cell carcinoma. *Asian pac. J. Cancer Prev.* 15 (5), 2309–2312. doi:10.7314/apjcp.2014.15.5.2309

Johnson, L. A., Vaidya, S. V., Goldfarb, R. H., and Mathew, P. A. (2003). 2B4(CD244)-mediated activation of NK cells reduces metastases of B16F10 melanoma in mice. *Anticancer Res.* 23 (5A), 3651–3655.

Ju, J. A., Godet, I., DiGiacomo, J. W., and Gilkes, D. M. (2020). RhoB is regulated by hypoxia and modulates metastasis in breast cancer. *Cancer Rep.* 3 (1), e1164. doi:10.1002/cnr2.1164

Kalluri, R., and Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* 119 (6), 1420–1428. doi:10.1172/JCI39104

Kannathasan, T., Kuo, W. W., Chen, M. C., Viswanadha, V. P., Shen, C. Y., Tu, C. C., et al. (2020). Chemoresistance-associated silencing of miR-4454 promotes colorectal cancer aggression through the GNL3L and NF-κB pathway. *Cancers* 12 (5), E1231. doi:10.3390/cancers12051231

Kingma, D. P., and Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv [cs.LG].

Kitamura, M., Wada, N., Nagata, S., Iizuka, N., Jin, Y. F., Tomoeda, M., et al. (2010). Malignant peripheral nerve sheath tumor associated with neurofibromatosis type 1, with metastasis to the heart: A case report. *Diagn. Pathol.* 5 (1), 2. doi:10.1186/1746-1596-5-2

Knouse, P., Hancock, C., Iwaz, S., and Kaiser, P. (2019). Metastatic carcinomatosis Cirrhosis: A rare pattern of metastasis. *Cureus* 11 (1), e3876. doi:10.7759/cureus.3876

Ku, S. Y., Rosario, S., Wang, Y., Mu, P., Seshadri, M., Goodrich, Z. W., et al. (2017). Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* 355 (6320), 78–83. doi:10.1126/science.aah4199

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44 (W1), W90–W97. doi:10.1093/nar/gkw377

Kumar, S., Oien, D. B., Khurana, A., Cliby, W., Hartmann, L., Chien, J., et al. (2019). Coiled-coil and C2 domain-containing protein 1A (CC2D1A) promotes

chemotherapy resistance in ovarian cancer. *Front. Oncol.* 9, 986. doi:10.3389/fonc.2019.00986

Li, C., Ge, M., Chen, D., Sun, T., Jiang, H., Xie, Y., et al. (2020). RPL21 siRNA blocks proliferation in pancreatic cancer cells by inhibiting DNA replication and inducing G1 arrest and apoptosis. *Front. Oncol.* 10, 1730. doi:10.3389/fonc.2020.01730

Li, F., Zhang, C., and Fu, L. (2019). PRR14 overexpression promotes cell growth, epithelial to mesenchymal transition and metastasis of colon cancer via the AKT pathway. *PLoS One* 14 (10), e0218839. doi:10.1371/journal.pone.0218839

Li, H., Cao, Y., Ma, J., Luo, L., and Ma, B. (2021). Expression and prognosis analysis of GINS subunits in human breast cancer. *Medicine* 100 (11), e24827. doi:10.1097/MD.0000000000024827

Li, M., Jin, X., Li, H., Wu, G., Wang, S., Yang, C., et al. (2020). Key genes with prognostic values in suppression of osteosarcoma metastasis using comprehensive analysis. *BMC Cancer* 20 (1), 65. doi:10.1186/s12885-020-6542-z

Li, M., Wang, L., Zhan, Y., Zeng, T., Zhang, X., Guan, X. Y., et al. (2019). Membrane metalloendopeptidase (MME) suppresses metastasis of esophageal squamous cell carcinoma (ESCC) by inhibiting FAK-RhoA signaling Axis. *Am. J. Pathol.* 189 (7), 1462–1472. doi:10.1016/j.ajpath.2019.04.007

Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16 (6), 321–332. doi:10.1038/nrg3920

Ling, L.-J., Lu, C., Zhou, G. P., and Wang, S. (2010). Ectopic expression of RhoBTB2 inhibits migration and invasion of human breast cancer cells. *Cancer Biol. Ther.* 10 (11), 1115–1122. doi:10.4161/cbt.10.11.13431

Liu, H., Liu, M., You, H., Li, X., and Li, X. (2020). Oncogenic network and hub genes for natural killer/T-cell lymphoma utilizing WGCNA. *Front. Oncol.* 10, 223. doi:10.3389/fonc.2020.00223

Liu, H., Zhou, Y., Qiu, H., Zhuang, R., Han, Y., Liu, X., et al. (2021). Rab26 suppresses migration and invasion of breast cancer cells through mediating autophagic degradation of phosphorylated Src. *Cell Death Dis.* 12 (4), 284. doi:10.1038/s41419-021-03561-7

Liu, J., Wang, D., Zhang, C., Zhang, Z., Chen, X., Lian, J., et al. (2018). Identification of liver metastasis-associated genes in human colon carcinoma by mRNA profiling. *Chin. J. Cancer Res.* 30 (6), 633–646. doi:10.21147/j.issn.1000-9604.2018.06.08

Liu, K., Tang, Z., Huang, A., Chen, P., Liu, P., Yang, J., et al. (2017). Glyceraldehyde-3-phosphate dehydrogenase promotes cancer growth and metastasis through upregulation of SNAIL expression. *Int. J. Oncol.* 50 (1), 252–262. doi:10.3892/ijo.2016.3774

Liu, Y., Wang, D., Lei, M., Gao, J., Cui, Y., Jin, X., et al. (2021). GABARAP suppresses EMT and breast cancer progression via the AKT/mTOR signaling pathway. *Aging* 13 (4), 5858–5874. doi:10.18632/aging.202510

Lopez-Charcas, O., Espinosa, A. M., Alfaro, A., Herrera-Carrillo, Z., Ramirez-Cordero, B. E., Cortes-Reynosa, P., et al. (2018). The invasiveness of human cervical cancer associated to the function of NaV1.6 channels is mediated by MMP-2 activity. *Sci. Rep.* 8 (1), 12995. doi:10.1038/s41598-018-31364-y

Luger, D., Yang, Y. A., Raviv, A., Weinberg, D., Banerjee, S., Lee, M. J., et al. (2013). Expression of the B-cell receptor component CD79a on immature myeloid cells contributes to their tumor promoting effects. *PLoS One* 8 (10), e76115. doi:10.1371/journal.pone.0076115

Luu, X., Leonard, S., and Joseph, K.-A. (2015). Dermatomyositis presenting as a paraneoplastic syndrome with resolution of symptoms following surgical management of underlying breast malignancy. *J. Surg. Case Rep.* 2015 (7), rjv075. doi:10.1093/jscr/rjv075

Lv, Y., Wang, X., Li, X., Xu, G., Bai, Y., Wu, J., et al. (2020). Nucleotide de novo synthesis increases breast cancer stemness and metastasis via cGMP-PKG-MAPK signaling pathway. *PLoS Biol.* 18 (11), e3000872. doi:10.1371/journal.pbio.3000872

Lyons, S. M., Alizadeh, E., Mannheimer, J., Schuamberg, K., Castle, J., Schroder, B., et al. (2016). Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas. *Biol. Open* 5 (3), 289–299. doi:10.1242/bio.013409

Madreiter-Sokolowski, C. T., Gottschalk, B., Sokolowski, A. A., Malli, R., and Graier, W. F. (2021). Dynamic control of mitochondrial Ca^{2+} levels as a survival strategy of cancer cells. *Front. Cell Dev. Biol.* 9, 614668. doi:10.3389/fcell.2021.614668

Mao-De, L., and Jing, X. (2007). Ribosomal proteins and colorectal cancer. *Curr. Genomics* 8 (1), 43–49. doi:10.2174/138920207780076938

Marsh, T., and Debnath, J. (2020). Autophagy suppresses breast cancer metastasis by degrading NBR1. *Autophagy* 16 (6), 1164–1165. doi:10.1080/15548627.2020.1753001

Martorell-Marugán, J., et al. (2019). "Deep learning in omics data analysis and precision medicine," in *Computational biology*. Editor H. Husi (Brisbane (AU): Codon Publications).

McLaren, S., and University of Western Australia, S. (2009). *School of,* mitochondrial ATP Synthase and associated bioenergetic proteins: Can expression predict Response to Chemotherapy in locally advanced breast cancer? Australia: University Of Western Australia, 382.

Medina-Ramirez, C. M., Goswami, S., Smirnova, T., Bamira, D., Benson, B., Ferrick, N., et al. (2011). Apoptosis inhibitor ARC promotes breast tumorigenesis, metastasis, and chemoresistance. *Cancer Res.* 71 (24), 7705–7715. doi:10.1158/0008-5472.CAN-11-2192

Milkiewicz, P., Mutimer, D., Hubscher, S. G., and Elias, E. (1999). Autoimmune liver disease in patients with neoplastic diseases. *Eur. J. Gastroenterol. Hepatol.* 11 (5), 569–573. doi:10.1097/00042737-199905000-00018

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *J. Big Data* 2 (1), 1–21. doi:10.1186/s40537-014-0007-7

Nałęcz, K. A. (2020). Amino acid transporter SLC6A14 (ATB0, +) - a target in combined anti-cancer therapy. *Front. Cell Dev. Biol.* 8, 594464. doi:10.3389/fcell.2020.594464

Ohtaki, S., Wanibuchi, M., Kataoka-Sasaki, Y., Sasaki, M., Oka, S., Noshiro, S., et al. (2017). ACTC1 as an invasion and prognosis marker in glioma. *J. Neurosurg.* 126 (2), 467–475. doi:10.3171/2016.1.JNS152075

Okusha, Y., Eguchi, T., Tran, M. T., Sogawa, C., Yoshida, K., Itagaki, M., et al. (2020). Extracellular vesicles enriched with moonlighting metalloproteinase are highly transmissive, pro-tumorigenic, and trans-activates cellular communication network factor (CCN2/CTGF): CRISPR against cancer. *Cancers* 12 (4), E881. doi:10.3390/cancers12040881

Paschen, A., Sucker, A., Hill, B., Moll, I., Zapatka, M., Nguyen, X. D., et al. (2009). Differential clinical significance of individual NKG2D ligands in melanoma: Soluble ULBP2 as an indicator of poor prognosis superior to S100B. *Clin. Cancer Res.* 15 (16), 5208–5215. doi:10.1158/1078-0432.CCR-09-0886

Pelosof, L. C., and Gerber, D. E. (2010). Paraneoplastic syndromes: An approach to diagnosis and treatment. *Mayo Clin. Proc.* 85 (9), 838–854. doi:10.4065/mcp.2010.0099

Pitarresi, J. R., Norgard, R. J., Chiarella, A. M., Suzuki, K., Bakir, B., Sahu, V., et al. (2021). PTHrP drives pancreatic cancer growth and metastasis and reveals a new therapeutic vulnerability. *Cancer Discov.* 11 (7), 1774–1791. doi:10.1158/2159-8290.CD-20-1098

Plantone, D., Renna, R., Sbardella, E., and Koudriavtseva, T. (2015). Concurrence of multiple sclerosis and brain tumors. *Front. Neurol.* 6, 40. doi:10.3389/fneur.2015.00040

Racanelli, V., Prete, M., Minoia, C., Favoino, E., and Perosa, F. (2008). Rheumatic disorders as paraneoplastic syndromes. *Autoimmun. Rev.* 7 (5), 352–358. doi:10.1016/j.autrev.2008.02.001

Rae, D. T., Hocum, J. D., Bii, V., Deeg, H. J., and Trobridge, G. D. (2015). A novel retroviral mutagenesis screen identifies prognostic genes in RUNX1 mediated myeloid leukemogenesis. *Oncotarget* 6 (31), 30664–30674. doi:10.18632/oncotarget.5133

Roy, L. D., Ghosh, S., Pathangey, L. B., Tinder, T. L., Gruber, H. E., and Mukherjee, P. (2011). Collagen induced arthritis increases secondary metastasis in MMTV-PyV MT mouse model of mammary cancer. *BMC Cancer* 11, 365. doi:10.1186/1471-2407-11-365

Scheau, C., Badarau, I. A., Costache, R., Caruntu, C., Mihai, G. L., Didilescu, A. C., et al. (2019). The role of matrix metalloproteinases in the epithelial-mesenchymal transition of hepatocellular carcinoma. *Anal. Cell. Pathol.* 2019, 9423907. doi:10.1155/2019/9423907

Schulten, H.-J., Bangash, M., Karim, S., Dallol, A., Hussein, D., Merdad, A., et al. (2017). Comprehensive molecular biomarker identification in breast cancer brain metastases. *J. Transl. Med.* 15 (1), 269. doi:10.1186/s12967-017-1370-x

Seetharaman, S., Flemyng, E., Shen, J., Conte, M. R., and Ridley, A. J. (2016). The RNA-binding protein LARP4 regulates cancer cell migration and invasion. *Cytoskeleton* 73 (11), 680–690. doi:10.1002/cm.21336

Seguin, L., Chaffanet, M., Sabatier, R., Jose, A., Garnier, S., Carbuccia, N., et al. (2018). A major response to carboplatin in a metastatic triple-negative breast cancer patient with somatic mutation of BRCA1 and RAD51B: When chemotherapy meets precision medicine. *Ann. Oncol.* 29, vi29. doi:10.1093/annonc/mdy314.030

Shah, A. A., and Casciola-Rosen, L. (2015). Cancer and scleroderma: A paraneoplastic disease with implications for malignancy screening. *Curr. Opin. Rheumatol.* 27 (6), 563–570. doi:10.1097/BOR.0000000000000222

Sharifi-Noghabi, H., Liu, Y., Erho, N., Shrestha, R., Alshalalfa, M., Davicioni, E., et al. *Deep Genomic Signature for early metastasis prediction in prostate cancer*. bioRxiv.

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). "Learning important features through propagating activation differences," in Proceedings of the 34th International Conference on Machine Learning-Volume 70, Sydney NSW Australia, August 6 - 11, 2017.

Sinha, P., Okoro, C., Foell, D., Freeze, H. H., Ostrand-Rosenberg, S., and Srikrishna, G. (2008). Proinflammatory S100 proteins regulate the accumulation of myeloid-derived suppressor cells. *J. Immunol.* 181 (7), 4666–4675. doi:10.4049/jimmunol.181.7.4666

Song, S., Jacobson, K. N., McDermott, K. M., Reddy, S. P., Cress, A. E., Tang, H., et al. (2016). ATP promotes cell survival via regulation of cytosolic [Ca2+] and Bcl-2/Bax ratio in lung cancer cells. *Am. J. Physiol. Cell Physiol.* 310 (2), C99–C114. doi:10.1152/ajpcell.00092.2015

Song, W., He, D., Chen, Y., Yeh, C. R., Hsu, I., Huang, Q., et al. (2018). Targeting newly identified ERβ/TGF-β1/SMAD3 signals with the FDA-approved anti-estrogen Faslodex or an ERβ selective antagonist in renal cell carcinoma. *Mol. Oncol.* 12 (12), 2055–2071. doi:10.1002/1878-0261.12377

Strömvall, K., Sundkvist, K., Ljungberg, B., Halin Bergstrom, S., and Bergh, A. (2017). Reduced number of CD169+ macrophages in pre-metastatic regional lymph nodes is associated with subsequent metastatic disease in an animal model and with poor outcome in prostate cancer patients. *Prostate* 77 (15), 1468–1477. doi:10.1002/pros.23407

Sun, Y., Zhou, Y., Xia, J., Wen, M., Wang, X., Zhang, J., et al. (2021). Abnormally high HIP1 expression is associated with metastatic behaviors and poor prognosis in ESCC. *Oncol. Lett.* 21 (2), 79. doi:10.3892/ol.2020.12340

Svozil, D., Kvasnicka, V., and Pospichal, J. í. (1997). Introduction to multi-layer feed-forward neural networks. *Chemom. Intelligent Laboratory Syst.* 39 (1), 43–62. doi:10.1016/s0169-7439(97)00061-0

Wang, M., Liu, Y., Qian, X., Wei, N., Tang, Y., and Yang, J. (2018). Downregulation of occludin affects the proliferation, apoptosis and metastatic properties of human lung carcinoma. *Oncol. Rep.* 40 (1), 454–462. doi:10.3892/or.2018.6408

Wang, X., Wang, L., Xu, Y., Zhang, G., Wu, Y., and Chen, P. (2018). PIAS1 inhibited the metastasis of gastric cancer cell by epithelial-mesenchymal transition regulation within the inflammatory microenvironment. *Oncol. Lett.* 15 (3), 3828–3837. doi:10.3892/ol.2018.7811

Wang, Y., Wu, Y., Xiao, K., Zhao, Y., Lv, G., Xu, S., et al. (2020). RPS24c isoform facilitates tumor angiogenesis via promoting the stability of MVIH in colorectal cancer. *Curr. Mol. Med.* 20 (5), 388–395. doi:10.2174/1566524019666191203123943

Wang, Z., Xiong, S., Mao, Y., Chen, M., Ma, X., Zhou, X., et al. (2016). Periostin promotes immunosuppressive premetastatic niche formation to facilitate breast tumour metastasis. *J. Pathol.* 239 (4), 484–495. doi:10.1002/path.4747

Wen, J., Min, X., Shen, M., Hua, Q., Han, Y., Zhao, L., et al. (2019). ACLY facilitates colon cancer cell metastasis by CTNNB1. *J. Exp. Clin. Cancer Res.* 38 (1), 401. doi:10.1186/s13046-019-1391-9

Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* 11 (1), 5120. doi:10.1038/s41467-020-18794-x

Wong, H. Y., Wang, G. M., Croessmann, S., Zabransky, D. J., Chu, D., Garay, J. P., et al. (2015). TMSB4Y is a candidate tumor suppressor on the Y chromosome and is deleted in male breast cancer. *Oncotarget* 6 (42), 44927–44940. doi:10.18632/oncotarget.6743

Wu, J., Tian, B., Yang, J., Huo, H., Song, Z., Yu, J., et al. (2020). Reduction of Hip2 suppresses gastric cancer cell proliferation, migration, invasion and tumorigenesis. *Transl. Cancer Res.* 9 (2), 774–785. doi:10.21037/tcr.2019.12.12

Xu, L., Zhou, R., Yuan, L., Wang, S., Li, X., Ma, H., et al. (2017). IGF1/IGF1R/STAT3 signaling-inducible IFITM2 promotes gastric cancer growth and metastasis. *Cancer Lett.* 393, 76–85. doi:10.1016/j.canlet.2017.02.014

Xu, L., Ziegelbauer, J., Wang, R., Wu, W. W., Shen, R. F., Juhl, H., et al. (2016). Distinct profiles for mitochondrial t-RNAs and small nucleolar RNAs in locally invasive and metastatic colorectal cancer. *Clin. Cancer Res.* 22 (3), 773–784. doi:10.1158/1078-0432.CCR-15-0737

Xu, Q., Liu, X., Chen, W., and Zhang, Z. (2010). Inhibiting adenoid cystic carcinoma cells growth and metastasis by blocking the expression of ADAM 10 using RNA interference. *J. Transl. Med.* 8 (1), 136. doi:10.1186/1479-5876-8-136

Xu, Y., Cui, X., and Wang, Y. (2021). Pan-cancer metastasis prediction based on Graph deep learning method. *Front. Cell Dev. Biol.* 9, 675978. doi:10.3389/fcell.2021.675978

Yamaguchi, M., Osuka, S., Weitzmann, M. N., Shoji, M., and Murata, T. (2016). Increased regucalcin gene expression extends survival in breast cancer patients: Overexpression of regucalcin suppresses the proliferation and metastatic bone activity in MDA-MB-231 human breast cancer cells *in vitro*. *Int. J. Oncol.* 49 (2), 812–822. doi:10.3892/ijo.2016.3538

Yang, R., Gaidamakov, S. A., Xie, J., Lee, J., Martino, L., Kozlov, G., et al. (2011). La-related protein 4 binds poly(A), interacts with the poly(A)-binding protein MLLE domain via a variant PAM2w motif, and can promote mRNA stability. *Mol. Cell. Biol.* 31 (3), 542–556. doi:10.1128/MCB.01162-10

Yano, Y., Akiba, J., Naito, Y., Sadashima, E., Cho, H., Hishima, T., et al. (2021). Sulfite oxidase is a novel prognostic biomarker of advanced gastric cancer. *Vivo* 35 (1), 229–237. doi:10.21873/invivo.12251

Yi, X., Luk, J. M., Lee, N. P., Peng, J., Leng, X., Guan, X. Y., et al. (2008). Association of mortalin (HSPA9) with liver cancer metastasis and prediction for early tumor recurrence. *Mol. Cell. Proteomics* 7 (2), 315–325. doi:10.1074/mcp.M700116-MCP200

Zeng, H., Yuan, F., Mi, Y., Xian, G., Qin, C., and Zhang, D. (2018). As an independent prognostic factor, USP6 promotes the invasion and metastasis of colon cancer. *Biochem. Biophys. Res. Commun.* 505 (3), 816–822. doi:10.1016/j.bbrc.2018.08.168

Zhang, B., Li, Y., Wu, Q., Xie, L., Barwick, B., Fu, C., et al. (2021). Acetylation of KLF5 maintains EMT and tumorigenicity to cause chemoresistant bone metastasis in prostate cancer. *Nat. Commun.* 12 (1), 1714. doi:10.1038/s41467-021-21976-w

Zhang, C., Li, H., Gao, J., Cui, X., Yang, S., and Liu, Z. (2021). Prognostic significance of ANO1 expression in cancers. *Medicine* 100 (4), e24525. doi:10.1097/MD.0000000000024525

Zhang, F., Ying, L., Jin, J., Feng, J., Chen, K., Huang, M., et al. (2018). GAP43, a novel metastasis promoter in non-small cell lung cancer. *J. Transl. Med.* 16 (1), 310. doi:10.1186/s12967-018-1682-5

Zhang, J., Huang, J. Z., Zhang, Y. Q., Zhang, X., Zhao, L. Y., Li, C. G., et al. (2020). Microtubule associated protein 9 inhibits liver tumorigenesis by suppressing ERCC3. *EBioMedicine* 53, 102701. doi:10.1016/j.ebiom.2020.102701

Zhang, Y., Xia, M., Jin, K., Wang, S., Wei, H., Fan, C., et al. (2018). Function of the c-Met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities. *Mol. Cancer* 17 (1), 45. doi:10.1186/s12943-018-0796-y

Zheng, G., Ma, Y., Zou, Y., Yin, A., Li, W., and Dong, D. (2018). Hcmdb: The human cancer metastasis database. *Nucleic Acids Res.* 46 (D1), D950–D955. doi:10.1093/nar/gkx1008

Zhou, D., Tang, W., Zhang, Y., and An, H. X. (2019). JAM3 functions as a novel tumor suppressor and is inactivated by DNA methylation in colorectal cancer. *Cancer Manag. Res.* 11, 2457–2470. doi:10.2147/CMAR.S189937

Zhou, Y., Zang, Y., Yang, Y., Xiang, J., and Chen, Z. (2019). Candidate genes involved in metastasis of colon cancer identified by integrated analysis. *Cancer Med.* 8 (5), 2338–2347. doi:10.1002/cam4.2071

frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 25 July 2022
DOI 10.3389/fphys.2022.978222

Check for updates

OPEN ACCESS

EDITED BY
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

REVIEWED BY
Shumao Pang,
Guangzhou Medical University, China
Yao Wu,
Children's National Hospital, United States

*CORRESPONDENCE
Shasha Du,
duss0202@163.com
Jun Cheng,
chengjun583@qq.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

RECEIVED 25 June 2022
ACCEPTED 05 July 2022
PUBLISHED 25 July 2022

CITATION
Tan P, Huang W, Wang L, Deng G, Yuan Y, Qiu S, Ni D, Du S and Cheng J (2022), Deep learning predicts immune checkpoint inhibitor-related pneumonitis from pretreatment computed tomography images. *Front. Physiol.* 13:978222. doi: 10.3389/fphys.2022.978222

COPYRIGHT
© 2022 Tan, Huang, Wang, Deng, Yuan, Qiu, Ni, Du and Cheng. 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.

Deep learning predicts immune checkpoint inhibitor-related pneumonitis from pretreatment computed tomography images

Peixin Tan[1†], Wei Huang[1†], Lingling Wang[2,3,4†], Guanhua Deng[5], Ye Yuan[2,3,4], Shili Qiu[2,3,4], Dong Ni[2,3,4], Shasha Du[1*] and Jun Cheng[2,3,4*]

[1]Department of Radiation Oncology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [2]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, [3]Medical Ultrasound Image Computing (MUSIC) Laboratory, Shenzhen University, Shenzhen, China, [4]Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, China, [5]Department of Oncology, Guangdong Sanjiu Brain Hospital, Guangzhou, China

Immune checkpoint inhibitors (ICIs) have revolutionized the treatment of lung cancer, including both non-small cell lung cancer and small cell lung cancer. Despite the promising results of immunotherapies, ICI-related pneumonitis (ICIP) is a potentially fatal adverse event. Therefore, early detection of patients at risk for developing ICIP before the initiation of immunotherapy is critical for alleviating future complications with early interventions and improving treatment outcomes. In this study, we present the first reported work that explores the potential of deep learning to predict patients who are at risk for developing ICIP. To this end, we collected the pretreatment baseline CT images and clinical information of 24 patients who developed ICIP after immunotherapy and 24 control patients who did not. A multimodal deep learning model was constructed based on 3D CT images and clinical data. To enhance performance, we employed two-stage transfer learning by pre-training the model sequentially on a large natural image dataset and a large CT image dataset, as well as transfer learning. Extensive experiments were conducted to verify the effectiveness of the key components used in our method. Using five-fold cross-validation, our method accurately distinguished ICIP patients from non-ICIP patients, with area under the receiver operating characteristic curve of 0.918 and accuracy of 0.920. This study demonstrates the promising potential of deep learning to identify patients at risk for developing ICIP. The proposed deep learning model enables efficient risk stratification, close monitoring, and prompt management of ICIP, ultimately leading to better treatment outcomes.

KEYWORDS

immune checkpoint inhibitor-related pneumonitis, deep learning, transfer learning, contrastive learning, CT images, lung cancer

1 Introduction

Since the first immune checkpoint inhibitor (ICI) ipilimumab was approved by the Food and Drug Administration for treating melanoma in 2011, ICIs have become standard treatments for many cancers such as lung cancer, renal cell carcinoma, Hodgkin lymphoma, and hepatocellular carcinoma (Akinleye and Rasool, 2019; Xin Yu et al., 2019; Robert, 2020; Vaddepally et al., 2020). Although ICIs produce remarkable immune response by immune upregulation and demonstrate improved cancer-related outcomes, they induce a unique spectrum of toxicities, called immune-related adverse events (irAEs) (Martins et al., 2019; Ramos-Casals et al., 2020). These irAEs can occur in multiple organ systems, where uncontrolled immune response is generated against healthy tissue. Due to different organ systems affected, there are various types of irAEs, including dermatitis, encephalitis, uveitis, hepatitis, and pneumonitis. Among these, immune checkpoint inhibitor-related pneumonitis (ICIP) is one of the most concerned adverse events because it is potentially life-threatening (Naidoo et al., 2017).

Lung cancer is the second most common cancer and the leading cause of cancer death worldwide. ICIs have shown significant clinical benefit in the treatment of advanced non-small cell lung cancer (NSCLC) (Reck et al., 2016). The incidence of ICIP in NSCLC is 4.1% as reported in a prospective study, while some real-word studies outside of clinical trials report a much higher incidence, ranging from 7% to 19% (Sears et al., 2019; Atchley et al., 2021). The time to onset of ICIP can vary from 9 days to 24.3 months after the initiation of immunotherapy (Nishino et al., 2015; Naidoo et al., 2017), with a median time of 52.5 days (Atchley et al., 2021). Most patients with ICIP have high severity that requires hospitalization, and about 27% of them die during the treatment for ICIP (Atchley et al., 2021). Unfortunately, the pathogenesis of ICIP has not been clearly elucidated. Possible risk factors include prior thoracic radiotherapy, pulmonary comorbidities, smoking status, and PD-1 inhibitors (Howell et al., 2015; Delaunay et al., 2017; Khunger et al., 2017; Pillai et al., 2018; Winer et al., 2018). However, it is challenging to accurately predict ICIP based on these clinical risk factors. In order to improve lung cancer treatment and outcomes, there is an urgent need for early prediction of ICIP, which enables risk stratification before starting immunotherapy and allows a close monitoring of high-risk patients during treatment.

Radiomics is a rapidly evolving research area in personalized precision medicine that aims to extract informative radiomic features from medical images and relate these features to clinical and biological endpoints. Computed tomography (CT) is routinely used for diagnosing lung cancer and assessing treatment response. CT-based quantitative radiomics approaches have been successfully applied to various tasks, such as lesion classification (Naidoo et al., 2017; Gitto et al., 2021; Xu et al., 2021), prediction of prognosis and treatment response (van Timmeren et al., 2017; Yang et al., 2019; Chetan and Gleeson, 2021), and genotype-phenotype associations (Rios Velazquez et al., 2017; Thawani et al., 2018; Zanfardino et al., 2019; Wu et al., 2021). There are very few studies focusing on the prediction of ICIP using radiomics. To the best of our knowledge, we only found two closely related ones. The first study reported a 100% accuracy of classification based on baseline chest CT images, but only two ICIP patients were enrolled (Colen et al., 2018). Mu et al. (2020) performed a radiomics analysis of PET/CT images to predict severe immune-related adverse events and achieved an area under the receiver operating characteristic curve (AUC) of 0.88 in a prospective validation cohort (Mu et al., 2020).

Predicting ICIP by conventional radiomics methods has two limitations regarding to the two steps in radiomics analysis pipeline. Radiomics requires first the segmentation of region of interest (ROI) and then the extraction of a fixed set of features from ROI. The first limitation is that it is unclear what region in pretreatment CT images should be used as ROI due to the lack of guidance for regional predilection of pneumonitis. The second limitation is that the predefined set of features may not be optimal for the final prediction task. Recent studies have demonstrated the excellent performance of deep learning models in computer-aided diagnosis (Yu et al., 2020; Zheng et al., 2020; Cheng et al., 2022b; Qian et al., 2022). Compared with hand-crafted features, deep learning models can directly learn discriminative features from images without prior segmentation of ROI and thus may provide a better prediction of ICIP.

In this study, we aim to develop a deep learning model based on clinical data and pretreatment chest CT images to predict the risk of ICIP in lung cancer patients. To this end, we collect a relatively large dataset consisting of ICIP and non-ICIP patients and propose a deep learning model in which multimodal data, two-stage transfer learning, and contrastive learning are used. Extensive experiments are conducted to assess the performance of different settings. The results demonstrate that the use of the aforementioned three strategies is effective and achieves state-of-the-art performance with an AUC of 0.918.

2 Materials and methods

2.1 Patients and data collection

This study was approved by the Ethics Committee of Guangdong Provincial People's Hospital, and the requirement for informed consent was waived. Figure 1 shows the detailed inclusion and exclusion criteria for preparing the patient cohort. A total of 353 lung cancer patients were treated with ICIs between January 2016 and December 2020 at our institute. We excluded 51 patients who received thoracic radiotherapy because radiotherapy can induce radiation pneumonitis which is

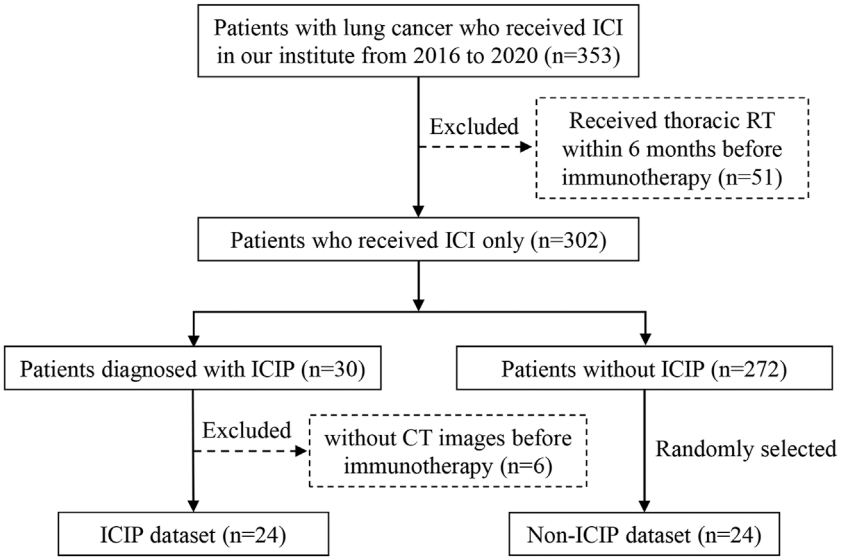


FIGURE 1
Flowchart for patient enrollment.

difficult to be distinguished from ICIP. Among the remaining patients, 30 of them developed ICIP, resulting in an incidence of 8.50% which is comparable to the data reported in previous studies (Sears et al., 2019; Atchley et al., 2021). We used the same criteria in a previous study to define ICIP (Cheng et al., 2022a). After excluding six patients who did not have CT scans before the start of immunotherapy, we finally got 24 patients for the ICIP dataset. To match the sample size of the ICIP dataset, we randomly chose 24 patients who did not develop ICIP to construct the control dataset, i.e., the non-ICIP dataset.

The collected chest CT images are within 6 months before the start of immunotherapy, which were produced by two different scanners, Philips iCT 256 and Philips ingenuity CT. Thoracic CT scans containing the whole lung were analyzed using a multi-slice helical technique at 120 kVp, mean exposure of 205 mAs, axial resolution of 5 mm, and mean in-plane resolution of 0.8174 mm.

2.2 Development of the deep learning model

2.2.1 Data preprocessing

For 3D CT scans, cropping, padding, and resizing techniques were used to convert the CT volume into a 192 × 192 × 224 matrix as the network input. We used a suitable window width for lung tissue from -500 to 1500 Hounsfield units (Zhang et al., 2020) to linearly rescale the pixel value to (0, 1) by the min-max method. Common data augmentation techniques including random flipping, noise, and affine transformation were used. For clinical information, as shown in Table 1, categorical variables were converted to distinct numbers so as to be input to models.

2.2.2 Network architecture

The overall network architecture is shown in Figure 2. To predict ICIP, we first built an image network and a clinical network based on pretreatment CT images and clinical data, respectively. Duo to the relatively small size of our dataset, a lightweight network, 3D ResNet18, was chosen as the backbone of the image network. Then, a multimodal fusion network was constructed by combining the nine clinical features (Table 1) and the image features learned from the image network. The clinical features and image features were fused by direct concatenation. Cross-entropy loss was used to supervise the ICIP prediction task. To enhance the prediction performance, two-stage transfer learning and contrastive learning strategies were used, which are introduced in the following sections.

2.2.3 Transfer learning

A two-stage transfer learning strategy inspired by (Altaf et al., 2021) was used to train our image network. We first downloaded the pre-trained model which was built using two massive natural image datasets (Kay et al., 2017; Monfort et al., 2020). The pre-trained weights may not be appropriate for our ICIP prediction task due to distributional shift between natural images and medical images. Therefore, in the second stage we fine-tuned the network using a large CT image dataset associated with

TABLE 1 Patient characteristics. *p* values less than 0.05 are highlighted with an asterisk.

Characteristic	ICIP dataset (n = 24)	Non-ICIP dataset (n = 24)	p value
Sex			0.022*
Female	0 (0.0%)	6 (25.0%)	
Male	24 (100.0%)	18 (75.0%)	
Age			0.261
Median	60	59	
Range	38–75	37–77	
Lesion location			0.9999
Upper left	9 (37.5%)	10 (41.7%)	
Upper right	9 (37.5%)	7 (29.2%)	
Lower left	3 (12.5%)	2 (8.3%)	
Lower right	3 (12.5%)	3 (12.5%)	
Mediastinal	0 (0.0%)	1 (4.2%)	
Middle right	0 (0.0%)	1 (4.2%)	
Histologic type			0.337
Adenocarcinoma	16 (66.7%)	16 (66.7%)	
Squamous cell carcinoma	7 (29.2%)	4 (16.7%)	
Adenosquamous carcinoma	0 (0.0%)	1 (4.2%)	
Small cell endocrine carcinoma	1 (4.2%)	0 (0.0%)	
Large cell endocrine carcinoma	0 (0.0%)	1 (4.2%)	
Lymphoepithelioma-like carcinoma	0 (0.0%)	2 (8.3%)	
T stage			0.496
T0	1 (4.2%)	1 (4.2%)	
T1	5 (20.8%)	2 (8.3%)	
T2	6 (25.0%)	11 (45.8%)	
T3	4 (16.7%)	2 (8.3%)	
T4	8 (33.3%)	8 (33.3%)	
N stage			0.233
N0	1 (4.2%)	2 (8.3%)	
N1	0 (0.0%)	0 (0.0%)	
N2	13 (54.2%)	7 (29.2%)	
N3	10 (41.7%)	15 (62.5%)	
M stage			0.461
M0	6 (25.0%)	3 (12.5%)	
M1	18 (75.0%)	21 (87.5%)	
Surgery before immunotherapy			0.666
Yes	4 (16.7%)	2 (8.3%)	
No	20 (83.3%)	22 (91.7%)	
Radiotherapy before immunotherapy			0.023*
Yes	8 (33.3%)	1 (4.2%)	
No	16 (66.7%)	23 (95.8%)	

pneumonia (CC-CCII dataset) (Zhang et al., 2020). After the transfer of knowledge from a related task, the domain gap between the source and target tasks was significantly reduced. The two-stage transfer learning flowchart is shown in Figure 3.

2.2.4 Contrastive learning

Besides transfer learning, contrastive learning was also adopted to further boost the performance of the image network. The key idea of contrastive learning is to learn an embedding space in which positive sample pairs stay close to each

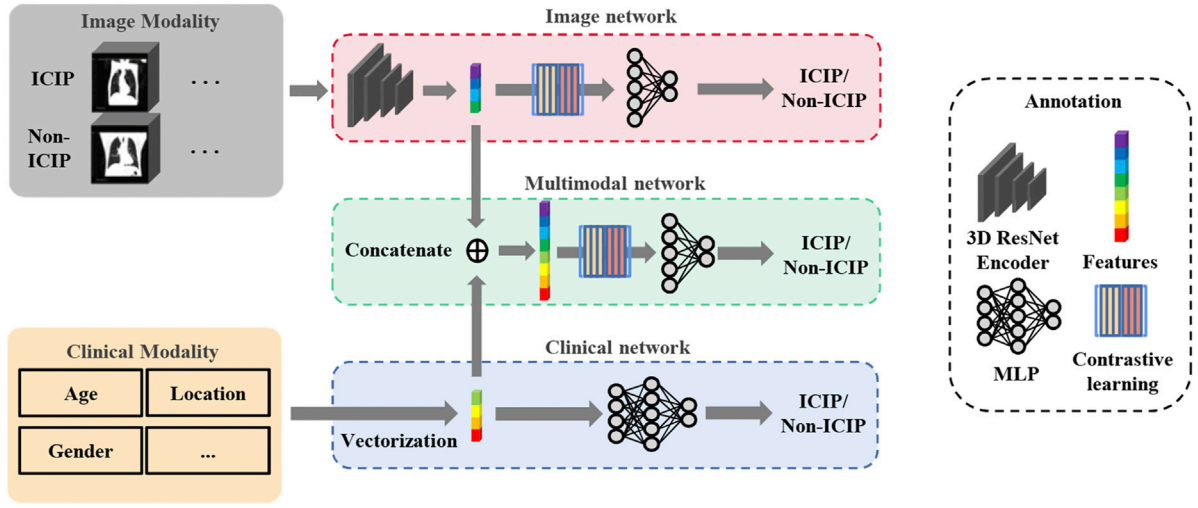


FIGURE 2
Overview of the network architecture for ICIP prediction. The top and bottom boxes show the image network and clinical network using CT images and clinical data, respectively, as input. The middle box represents the multimodal fusion network that combines image features and clinical features for ICIP prediction.

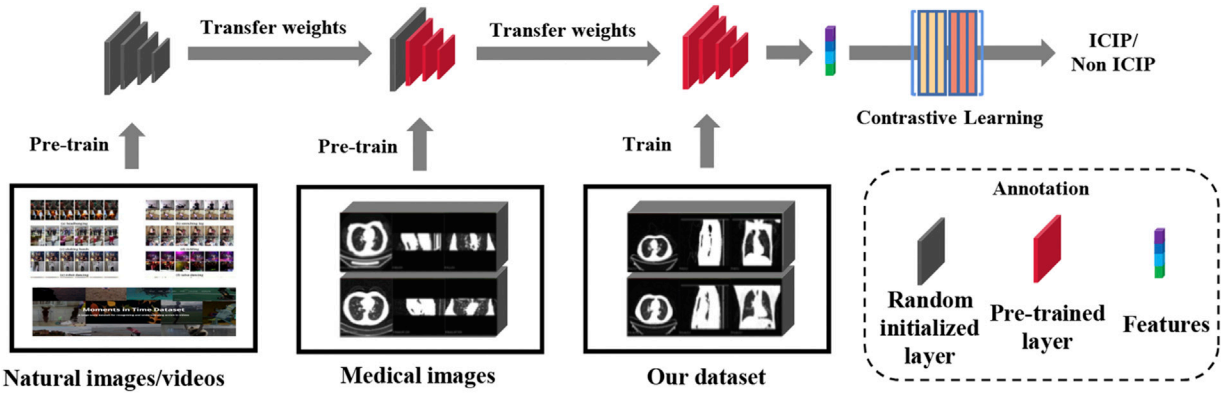


FIGURE 3
Flowchart of two-stage transfer learning.

other while negative ones are far apart. In essence, contrastive learning allows the model to learn high-level features about the data. The contrastive learning can be broken into three basic steps: sample pair construction, encoding, and loss minimization of representations. In our study, positive sample pairs were samples from the same class while negative sample pairs were samples from different classes. We then used 3D ResNet18 to encode the images as vector representations (Figure 2). Lastly, we maximized the similarity of the two vector representations of the positive sample pair and minimized that of the negative sample pair by minimizing a contrastive loss function. We took the cosine as the similarity metric. The contrastive loss function is defined by the following equations:

$$sim(x_1, x_2) = \frac{x_1 \cdot x_2}{\|x_1\| \cdot \|x_2\|}, \quad (1)$$

$$loss_{CL} = \sum_x sim(x, x_-) - \sum_x sim(x, x_+). \quad (2)$$

Here, sim represents the cosine similarity metric (x, x_) and (x, x_+) respectively denote the negative sample pair and positive sample pair. The overall loss function is defined by

$$loss_{all} = loss_{CE} + \gamma loss_{CL}, \quad (3)$$

where $loss_{CE}$ is the typical cross entropy loss, and γ is a hyper-parameter to control the weight of contrastive learning loss.

2.3 Implementation details

The proposed method was implemented using PyTorch on a workstation equipped with four NVIDIA RTX A6000 GPUs (48 GB memory each). In all comparative studies, we employed ResNet18 (He et al., 2016) as the backbone. Adam optimizer with the learning rate of 1e-4, 1e-4, and 5e-5 was employed to train the baseline model (without transfer learning), one-stage transfer learning model, and two-stage transfer learning model, respectively. The batch size and γ were set to 6 and 0.02, respectively, which were selected by grid search. Specifically, a finite number of values were tried, and the one with the best performance was selected. In all experiments, five-fold cross-validation was used, and the average performance was reported.

2.4 Statistical analysis

To compare the data distribution between the ICIP and non-ICIP datasets, Fisher's exact test was used for categorical variables, and Mann-Whitney U test was used for continuous variables. Two-tailed tests are used to determine significance at the 5% level. All statistical analyses were conducted using Statistical Product and Service Solutions (IBM SPSS, version 20.0).

To evaluate the classification performance, several typical metrics were used, including accuracy, sensitivity, specificity, precision, and F1-score. We considered ICIP as the positive class and non-ICIP as the negative class, so true positive (TP), false positive (FP), true negative (TN), and false negative (FN) can be accordingly defined. After getting the numbers of TP, FP, TN, and FN, the abovementioned performance metrics can be calculated using Eqs. 4–8. Since we used five-fold cross-validation, the average of these metrics were reported. We also used the area under the receiver operating characteristic (ROC) curve to evaluate model performance. Since every patient was tested for and only for once in five-fold cross-validation, we gathered the results across all the five folds, then plotted ROC curves, and calculated AUCs.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F1 - score = \frac{2Precision * Specificity}{Precision + Specificity} \tag{8}$$

3 Results

3.1 Patient characteristics

Among the 48 patients, there were 42 men and 6 women with an overall mean age of 58.00 years ± 9.75 (standard deviation). We collected nine clinical characteristics for the 48 patients. Table 1 shows these characteristics separately for the ICIP and non-ICIP datasets. Among the nine characteristics, there were significant differences between the two datasets for sex (Fishers' exact test p value = 0.022) and radiotherapy before immunotherapy (Fishers' exact test p value = 0.023), whereas no significant differences were observed for the remaining characteristics.

3.2 Performance of deep learning model to predict Immune checkpoint inhibitors-related pneumonitis

To explore and validate the effectiveness of the key components used in our method, we conducted extensive experiments. The comparison of quantitative performance is presented in Table 2. The details of different methods are provided as follows:

- Cli denotes the clinical network
- Im denotes the image network without using transfer learning and contrastive learning
- CI denotes the multimodal network built on both clinical data and CT images
- Im-1T denotes the image network with one-stage transfer learning
- Im-2T denotes the image network with two-stage transfer learning
- Im-2T-C denotes the image network with two-stage transfer learning and contrastive learning
- CI-2T denotes the multimodal network with two-stage transfer learning
- CI-2T-C denotes the multimodal network with two-stage transfer learning and contrastive learning

3.2.1 Effectiveness of multimodal data fusion

We first evaluated the effectiveness of combining images and clinical data to predict ICIP. To this end, we compared the classification performance of the multimodal network with that of the image network and the clinical network (CI vs. Cli and Im, Table 2). As shown in Table 2, the image network achieved an

TABLE 2 Quantitative analysis of key components in our method. The best results are highlighted in bold.

Method	AUC	Accuracy	Sensitivity	Specificity	Precision	F1-score
Cli	0.701	0.730 ± 0.045	0.660 ± 0.134	0.800 ± 0.200	0.817 ± 0.171	0.723 ± 0.041
Im	0.753	0.725 ± 0.075	0.740 ± 0.195	0.710 ± 0.175	0.728 ± 0.079	0.717 ± 0.081
CI	0.797	0.815 ± 0.078	0.790 ± 0.143	0.840 ± 0.089	0.837 ± 0.096	0.814 ± 0.079
Im-1T	0.821	0.830 ± 0.120	0.700 ± 0.200	0.960 ± 0.089	0.837 ± 0.096	0.824 ± 0.125
Im-2T	0.854	0.855 ± 0.087	0.800 ± 0.200	0.910 ± 0.125	0.920 ± 0.110	0.851 ± 0.090
Im-2T-C	0.901	0.920 ± 0.084	0.960 ± 0.089	0.880 ± 0.179	0.910 ± 0.131	0.918 ± 0.087
CI-2T	0.865	0.880 ± 0.130	0.920 ± 0.110	0.840 ± 0.167	0.860 ± 0.142	0.879 ± 0.131
CI-2T-C	0.918	0.920 ± 0.084	0.920 ± 0.110	0.920 ± 0.179	0.943 ± 0.128	0.918 ± 0.087

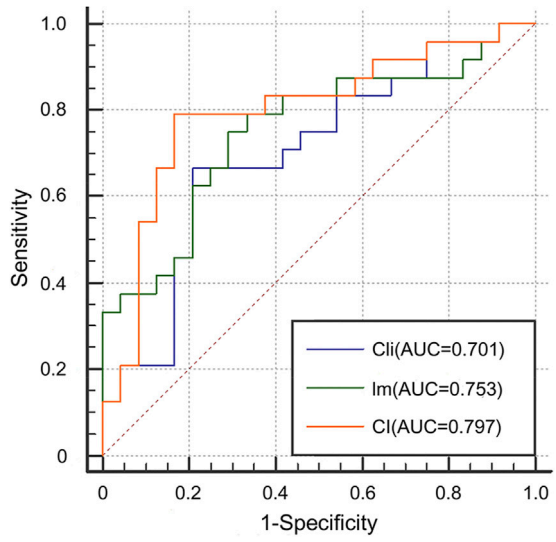


FIGURE 4
ROC curves for the Cli, Im, and CI networks to show the effectiveness of multimodal data fusion.

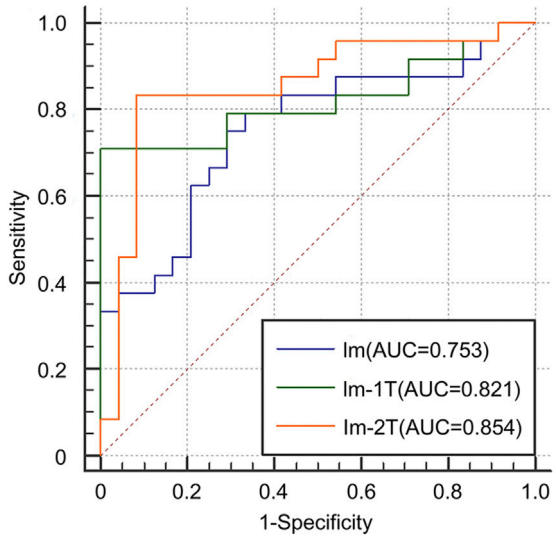


FIGURE 5
ROC curves for the Im, Im-1T, and Im-2T networks to show the effectiveness of two-stage transfer learning.

AUC of 0.753 (Im, Table 2), which was superior to the clinical network that yielded an AUC of 0.701 (Cli, Table 2). By utilizing both images and clinical data, the classification performance was significantly improved up to 0.797 (CI, Table 2). Other metrics in Table 2 also indicates that multimodal data fusion is beneficial to ICIP prediction. ROC curves are shown in Figure 4.

3.2.2 Effectiveness of two-stage transfer learning

We next evaluated the effectiveness of the proposed two-stage transfer learning strategy. To this end, we first used the image network without transfer learning as the baseline and then gradually incorporated one-stage and two-stage transfer learning. Compared with the baseline model trained from scratch, one-stage transfer learning brought large performance gain from 0.753 to 0.821 in term of AUC (Im vs. Im-1T, Table 2). Moreover, the use of the two-stage transfer learning further lifted the prediction performance. The AUC, accuracy, sensitivity, and specificity were 0.854, 0.855, 0.800, and 0.910, respectively (Im-2T, Table 2). The increasingly better performance from Im, Im-1T to Im-2T suggests that using a pre-trained network and fine-tuning on a large related dataset are essential to obtain good performance. ROC curves are shown in Figure 5.

3.2.3 Effectiveness of contrastive learning

Finally, we evaluated the effectiveness of the contrastive learning strategy in the image network and the multimodal network. The image network with two-stage transfer learning but without contrastive learning achieved AUC of 0.854, accuracy of 0.855, sensitivity of 0.800, and specificity of 0.910 (Im-2T, Table 2). Adding contrastive learning gave a boost in

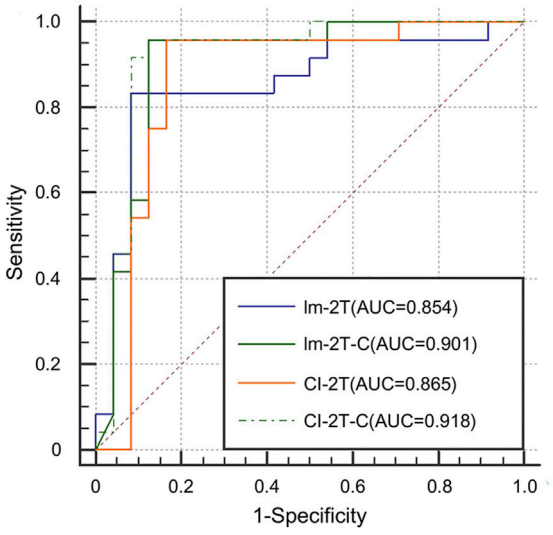


FIGURE 6
ROC curves for the Im-2T, Im-2T-C, CI-2T, and CI-2T-C networks to show the effectiveness of contrastive learning.

performance. The resulting AUC, accuracy, sensitivity, and specificity were 0.901, 0.920, 0.960, and 0.880, respectively (Im-2T-C, Table 2). Similarly, performance gain was also observed when incorporating contrastive learning into the multimodal network. The multimodal network with two-stage transfer learning yielded AUC of 0.865, accuracy of 0.880, sensitivity of 0.920, and specificity of 0.840 (CI-2T, Table 2). The use of contrastive learning increased the performance by a large margin. The resulting AUC, accuracy, sensitivity, and specificity were 0.918, 0.920, 0.920, and 0.920, respectively (CI-2T-C, Table 2). ROC curves are shown in Figure 6.

3.3 Visualization

To gain an understanding of why the performance was improved when the key components such as two-stage transfer learning were introduced to the model, the class activation maps were generated by the gradient of the deep learning to highlight the important regions within the input image (Gotkowski et al., 2021). As shown in Figure 7, the primary attention of the one-stage transfer learning model (Im-1T, first row) was not focused on the lung area. However, the use of two-stage transfer learning and especially contrastive learning brought more attention to the lung area (Im-2T and Im-2T-C, second and third rows). Interestingly, fusing clinical and image features made the network concentrate on the whole lung (CI-2T-C, fourth row), indicating that the whole lung is crucial and informative for ICIP prediction. This makes sense as ICIP can occur anywhere in lung.

4 Discussion

As ICI treatment is becoming more frequently used in lung cancer patients, an increasing number of irAEs (i.e., ICIP) are being reported. ICIP is potentially fatal. Thus, early prediction of ICIP is crucial for improving treatment outcomes. However, based on clinical factors or pretreatment CT images, it is very challenging for doctors to predict whether ICIP will occur prior to immunotherapy. Therefore, there is a critical need for an accurate and automated approach to assist doctors in identifying patients at risk for ICIP before immunotherapy, which allows personalized treatment options and reduces the number of deaths due to severe ICIP. In this study, we developed the first deep learning model for predicting ICIP using clinical information and pretreatment baseline chest CT images. In addition to the use of multimodal data, we also introduced two-stage transfer learning and contrastive learning in our model development. We evaluated our method using five-fold cross-validation on 24 ICIP patients and 24 non-ICIP patients. The results demonstrated that the deep learning model accurately differentiated between ICIP and non-ICIP patients, with an AUC of 0.918.

Few prior studies have demonstrated the utility of radiomics to predict irAEs. Colen et al. (2018) presented the first reported work exploring the potential of CT-based radiomics to predict patients at risk for developing ICIP and reported an AUC of 1 (Colen et al., 2018). Although the performance was extremely high, this study only included 2 ICIP cases and suffered from severe class imbalance problem. By contrast, our study used a balanced dataset consisting of 24 ICIP cases and 24 non-ICIP cases. Mu et al. proposed a PET/CT based radiomics approach to predict severe irAEs in patients with NSCLC (Mu et al., 2020). A total of 30 cases with severe irAEs and 164 control cases were curated in the patient cohorts. The radiomics approach yielded an AUC of 0.88 in the prospective validation cohort. However, this work is based on PET/CT which is not widely available in hospitals and thus has limited utility. In contrast to the traditional radiomics methods that extracted a fixed set of image features, our study proposed a deep learning model that can directly learn discriminative features from CT images and demonstrated a better performance with an AUC of 0.918.

The superiority of our method can be attributed to the use of multimodal data fusion, two-stage transfer learning, and contrastive learning in our deep learning model. The effectiveness of these key components was validated by extensive ablation studies. The multimodal data fusion model outperformed the models built on either clinical data or CT images by a large margin. This suggests that the two kinds of data harbor complementary information. Thus, the ICIP prediction task can greatly benefit from this fusion approach. Training deep learning models requires a large dataset. However, in medical applications, oftentimes, only a small dataset is available due to low incidence of disease or expensive cost of data collection. Our

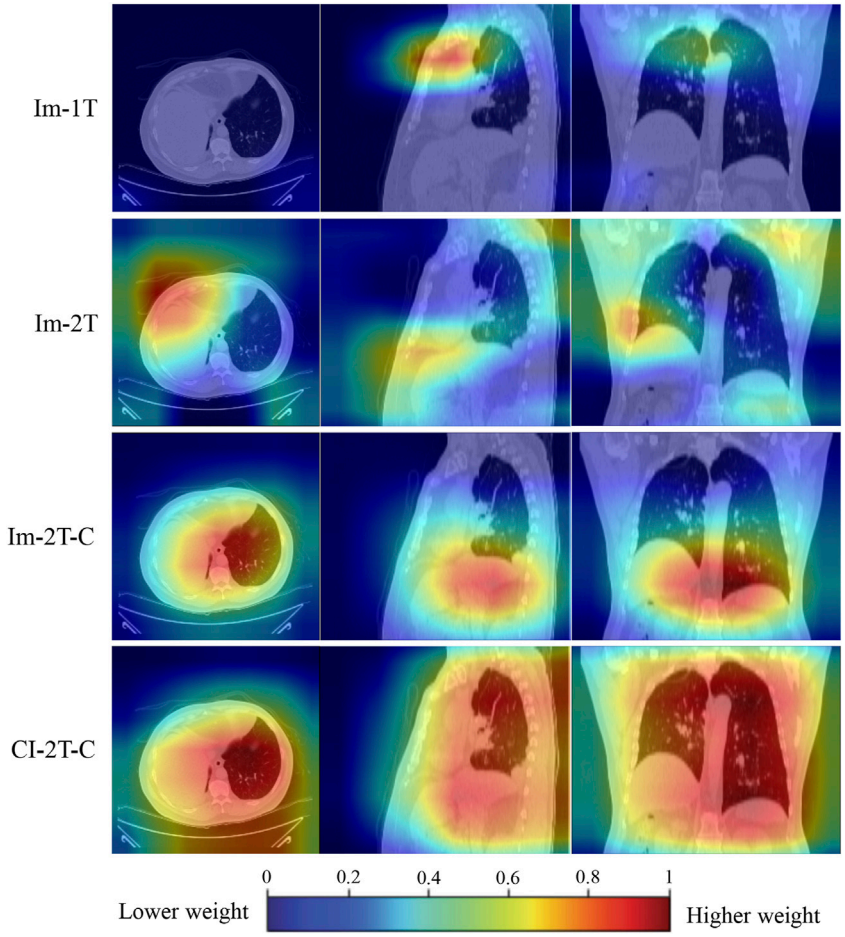


FIGURE 7
Activation maps for a non-ICIP sample to show the most "important" regions that different methods consider. The red color represents a higher weight (i.e., more attention is paid to this region).

results confirm that transfer learning is helpful in this case. Simply using a pre-trained model learned on a large unrelated dataset (one-stage transfer learning) or subsequently retraining the model on a large related dataset (two-stage transfer learning) can improve the performance of ICIP prediction markedly. Moreover, contrastive learning can further enhance the feature representation ability by contrasting similar (positive) and dissimilar (negative) samples.

This study has several limitations. First, although our method was rigorously validated by five-fold cross-validation, the data used in this study was collected from a single institution, future efforts will concentrate on validating the findings in a larger multi-institutional cohort. Second, to maintain a balance of sample size between classes, we randomly selected a portion of patients without ICIP to match the sample size of the ICIP dataset. There might be an issue with this strategy as it does not reflect a real-world class distribution. Third, duo to the retrospective nature of this study, it may be prone to biases from missing data and reliance on available medical documentation for review. Prospective studies are needed in the future.

In conclusion, patients who will develop ICIP have subtle changes at their pretreatment baseline CT scans that could not be identified by the naked eye but could be detected by quantitative analysis. Our study presents the first deep learning model based on clinical data and CT images to predict patients at risk for developing ICIP. This model can accurately predict ICIP patients with an AUC of 0.918, which enables efficient risk stratification, close monitoring, and prompt management of ICIP. This will potentially improve ICI treatment outcomes in patients with lung 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 authors.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Guangdong Provincial People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

PT and JC contributed to conception and design of the study. PT, LW, WH, and GD collected the data. LW, YY, and SQ analyzed and interpreted the data. LW and PT wrote the first draft of the manuscript. DN, SD, and JC supervised the project and revised the manuscript. All authors contributed to the manuscript and approved the submitted version.

Funding

This study was supported by National Natural Science Foundation of China (61901275, 62171290, 81972970, and 82172671), Guangzhou Science and Technology Plan Foundation (2021-02-01-04-1002-0017), Shenzhen University Startup Fund (2019131), Shenzhen-Hong Kong Joint Research Program (SGDX20201103095613036), Natural Science Foundation of Guangdong Province of China (2020A1515010186), GDPH Supporting Fund (KY0120220119), and GDPH Supporting Fund for NSFC Program (8190120260). The funding sources have no involvement in the study.

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

Akinleye, A., and Rasool, Z. (2019). Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. *J. Hematol. Oncol.* 12, 92. doi:10.1186/s13045-019-0779-5

Altaf, F., Islam, S. M. S., Janjua, N. K., and Akhtar, N. (2021). "Boosting deep transfer learning for COVID-19 classification," in Proceedings - International Conference on Image Processing, ICIP, Anchorage, AK, USA, September 19–22, 2021. (IEEE Computer Society), 210–214. doi:10.1109/ICIP42928.2021.9506646

Atchley, W. T., Alvarez, C., Saxena-Beem, S., Schwartz, T. A., Ishizawar, R. C., Patel, K. P., et al. (2021). Immune checkpoint inhibitor-related pneumonitis in lung cancer: real-world incidence, risk factors, and management practices across six health care centers in North Carolina. *Chest* 160, 731–742. doi:10.1016/j.chest.2021.02.032

Cheng, J., Pan, Y., Huang, W., Huang, K., Cui, Y., Hong, W., et al. (2022a). Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning. *Med. Phys.* 49, 1547–1558. doi:10.1002/mp.15451

Cheng, J., Wang, H., Li, R., Li, X., Zhou, X., Yang, X., et al. (2022b). A two-stage multiresolution neural network for automatic diagnosis of hepatic echinococcosis from ultrasound images: a multicenter study. *Med. Phys.* 49, 3199–3212. doi:10.1002/mp.15548

Chetan, M. R., and Gleeson, F. V. (2021). Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur. Radiol.* 31, 1049–1058. doi:10.1007/s00330-020-07141-9

Colen, R. R., Fujii, T., Bilen, M. A., Kotrotsou, A., Abrol, S., Hess, K. R., et al. (2018). Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. *Invest. New Drugs* 36, 601–607. doi:10.1007/s10637-017-0524-2

Delaunay, M., Cadranel, J., Lusque, A., Meyer, N., Gounaut, V., Moro-Sibilot, D., et al. (2017). Immune-checkpoint inhibitors associated with interstitial lung disease in cancer patients. *Eur. Respir. J.* 50, 1700050. doi:10.1183/13993003.00050-2017

Gitto, S., Cuocolo, R., Annovazzi, A., Anelli, V., Acquasanta, M., Cincotta, A., et al. (2021). CT radiomics-based machine learning classification of atypical cartilaginous tumours and appendicular chondrosarcomas. *EBioMedicine* 68, 103407. doi:10.1016/j.ebiom.2021.103407

Gotkowski, K., Gonzalez, C., Bucher, A., and Mukhopadhyay, A. (2021). "M3d-CAM: a PyTorch library to generate 3D attention maps for medical deep learning," in *Informatik aktuell* (Regensburg: Springer Science and Business Media Deutschland GmbH), 217–222. doi:10.1007/978-3-658-33198-6_52

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 770–778. doi:10.1109/CVPR.2016.90

Howell, M., Lee, R., Bowyer, S., Fusi, A., and Lorigan, P. (2015). Optimal management of immune-related toxicities associated with checkpoint inhibitors in lung cancer. *Lung Cancer* 88, 117–123. doi:10.1016/j.lungcan.2015.02.007

Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., et al. (2017). The kinetics human action video dataset. *arXiv Comput. Vis. Pattern Recognit.*

Khunger, M., Rakshit, S., Pasupuleti, V., Hernandez, A. V., Mazzone, P., Stevenson, J., et al. (2017). Incidence of pneumonitis with use of programmed death 1 and programmed death-ligand 1 inhibitors in non-small cell lung cancer: a systematic review and meta-analysis of trials. *Chest* 152, 271–281. doi:10.1016/j.chest.2017.04.177

Martins, F., Sofiya, L., Sykiotis, G. P., Lamine, F., Maillard, M., Fraga, M., et al. (2019). Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. *Nat. Rev. Clin. Oncol.* 16, 563–580. doi:10.1038/s41571-019-0218-0

Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Bargal, S. A., Yan, T., et al. (2020). Moments in time dataset: one million videos for event understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 502–508. doi:10.1109/TPAMI.2019.2901464

Mu, W., Tunali, I., Qi, J., Schabath, M. B., and Gillies, R. J. (2020). Radiomics of 18 F fluorodeoxyglucose PET/CT images predicts severe immune-related adverse events in patients with NSCLC. *Radiol. Artif. Intell.* 2, e190063. doi:10.1148/ryai.2019190063

Naidoo, J., Wang, X., Woo, K. M., Iyriboz, T., Halpenny, D., Cunningham, J., et al. (2017). Pneumonitis in patients treated with anti-programmed death-1/programmed death ligand 1 therapy. *J. Clin. Oncol.* 35, 709–717. doi:10.1200/JCO.2016.68.2005

Nishino, M., Sholl, L. M., Hatabu, H., Ramaiya, N. H., and Hodi, F. S. (2015). Anti–PD-1–Related pneumonitis during cancer immunotherapy. *N. Engl. J. Med.* 373, 288–290. doi:10.1056/nejmc1505197

Pillai, R. N., Behera, M., Owonikoko, T. K., Kamphorst, A. O., Pakkala, S., Belani, C. P., et al. (2018). Comparison of the toxicity profile of PD-1 versus PD-L1 inhibitors in non–small cell lung cancer: a systematic analysis of the literature. *Cancer* 124, 271–277. doi:10.1002/cncr.31043

Qian, J., Li, R., Yang, X., Huang, Y., Luo, M., Lin, Z., et al. (2022). Hasa: hybrid architecture search with aggregation strategy for echinococcosis classification and ovary segmentation in ultrasound images. *Expert Syst. Appl.* 202, 117242. doi:10.1016/J.ESWA.2022.117242

Ramos-Casals, M., Brahmer, J. R., Callahan, M. K., Flores-Chávez, A., Keegan, N., Khamashta, M. A., et al. (2020). Immune-related adverse events of checkpoint inhibitors. *Nat. Rev. Dis. Prim.* 6, 38. doi:10.1038/s41572-020-0160-6

Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., et al. (2016). Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer. *N. Engl. J. Med.* 375, 1823–1833. doi:10.1056/nejmoa1606774

Rios Velazquez, E., Parmar, C., Liu, Y., Coroller, T. P., Cruz, G., Stringfield, O., et al. (2017). Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res.* 77, 3922–3930. doi:10.1158/0008-5472.CAN-17-0122

Robert, C. (2020). A decade of immune-checkpoint inhibitors in cancer therapy. *Nat. Commun.* 11, 3801. doi:10.1038/s41467-020-17670-y

Sears, C. R., Rivera, M. P., Camus, P., Gaga, M., Garon, E. B., Gould, M. K., et al. (2019). Knowledge gaps and research priorities in immune checkpoint inhibitor–related pneumonitis an official American thoracic society research statement. *Am. J. Respir. Crit. Care Med.* 200, E31–E43. doi:10.1164/rccm.201906-1202ST

Thawani, R., McLane, M., Beig, N., Ghose, S., Prasanna, P., Velcheti, V., et al. (2018). Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer* 115, 34–41. doi:10.1016/j.lungcan.2017.10.015

Vaddepally, R. K., Kharel, P., Pandey, R., Garje, R., and Chandra, A. B. (2020). Review of indications of FDA-approved immune checkpoint inhibitors per NCCN guidelines with the level of evidence. *Cancers* 12, 738. doi:10.3390/cancers12030738

van Timmeren, J. E., Leijenaar, R. T. H., van Elmpt, W., Reymen, B., Oberije, C., Monshouwer, R., et al. (2017). Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. *Radiother. Oncol.* 123, 363–369. doi:10.1016/j.radonc.2017.04.016

Winer, A., Nicholas Bodor, J., and Borghaei, H. (2018). Identifying and managing the adverse effects of immune checkpoint blockade. *J. Thorac. Dis.* 10, S480–S489. doi:10.21037/jtd.2018.01.111

Wu, H., Wu, C., Zheng, H., Wang, L., Guan, W., Duan, S., et al. (2021). Radiogenomics of neuroblastoma in pediatric patients: CT-based radiomics signature in predicting MYCN amplification. *Eur. Radiol.* 31, 3080–3089. doi:10.1007/s00330-020-07246-1

Xin Yu, J., Hubbard-Lucey, V. M., and Tang, J. (2019). Immuno-oncology drug development goes global. *Nat. Rev. Drug Discov.* 18, 899–900. doi:10.1038/d41573-019-00167-9

Xu, Q. Q., Shan, W. L., Zhu, Y., Huang, C. C., Bao, S. Y., Guo, L. L., et al. (2021). Prediction efficacy of feature classification of solitary pulmonary nodules based on CT radiomics. *Eur. J. Radiol.* 139, 109667. doi:10.1016/j.ejrad.2021.109667

Yang, L., Yang, J., Zhou, X., Huang, L., Zhao, W., Wang, T., et al. (2019). Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur. Radiol.* 29, 2196–2206. doi:10.1007/s00330-018-5770-y

Yu, J., Deng, Y., Liu, T., Zhou, J., Jia, X., Xiao, T., et al. (2020). Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics. *Nat. Commun.* 11, 4807. doi:10.1038/s41467-020-18497-3

Zanfardino, M., Franzese, M., Pane, K., Cavaliere, C., Monti, S., Esposito, G., et al. (2019). Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. *J. Transl. Med.* 17, 337. doi:10.1186/s12967-019-2073-2

Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., et al. (2020). Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* 181, 1360. doi:10.1016/j.cell.2020.08.029

Zheng, X., Yao, Z., Huang, Y., Yu, Y., Wang, Y. Y., Liu, Y., et al. (2020). Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat. Commun.* 11, 1236. doi:10.1038/s41467-020-15027-z

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fgene.2022.934683

Check for updates

OPEN ACCESS

EDITED BY
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

REVIEWED BY
Xingchen Li,
Peking University People's Hospital, China
Lixia Xu,
Tianjin Huanhu Hospital, China

*CORRESPONDENCE
Li Zhang,
cmuzhangli@126.com
Sheng Han,
hansheng@cmu.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 02 May 2022
ACCEPTED 18 July 2022
PUBLISHED 11 August 2022

CITATION
Hu T, Wang Y, Wang X, Wang R, Song Y, Zhang L and Han S (2022), Construction and validation of an angiogenesis-related gene expression signature associated with clinical outcome and tumor immune microenvironment in glioma.
Front. Genet. 13:934683.
doi: 10.3389/fgene.2022.934683

COPYRIGHT
© 2022 Hu, Wang, Wang, Wang, Song, Zhang and Han. 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.

Construction and validation of an angiogenesis-related gene expression signature associated with clinical outcome and tumor immune microenvironment in glioma

Tianhao Hu[1], Yutao Wang[2], Xiaoliang Wang[1], Run Wang[3], Yifu Song[1], Li Zhang[1]* and Sheng Han[1]*

[1]Department of Neurosurgery, The First Hospital of China Medical University, Shenyang, China, [2]Department of Urology, The First Hospital of China Medical University, Shenyang, China, [3]Department of Neurosurgery, Huazhong University of Science and Technology Union Shenzhen Hospital, Shenzhen, China

Background: Glioma is the most prevalent malignant intracranial tumor. Many studies have shown that angiogenesis plays a crucial role in glioma tumorigenesis, metastasis, and prognosis. In this study, we conducted a comprehensive analysis of angiogenesis-related genes (ARGs) in glioma.

Methods: RNA-sequencing data of glioma patients were obtained from TCGA and CGGA databases. *Via* consensus clustering analysis, ARGs in the sequencing data were distinctly classified into two subgroups. We performed univariate Cox regression analysis to determine prognostic differentially expressed ARGs and least absolute shrinkage and selection operator Cox regression to construct a 14-ARG risk signature. The CIBERSORT algorithm was used to explore immune cell infiltration, and the ESTIMATE algorithm was applied to calculate immune and stromal scores.

Results: We found that the 14-ARG signature reflected the infiltration characteristics of different immune cells in the tumor immune microenvironment. Additionally, total tumor mutational burden increased significantly in the high-risk group. We combined the 14-ARG signature with patient clinicopathological data to construct a nomogram for predicting 1-, 3-, and 5-year overall survival with good accuracy. The predictive value of the prognostic model was verified in the CGGA cohort. *SPP1* was a potential biomarker of glioma risk and was involved in the proliferation, invasion, and angiogenesis of glioma cells.

Conclusion: In conclusion, we established and validated a novel ARG risk signature that independently predicted the clinical outcomes of glioma patients and was associated with the tumor immune microenvironment.

KEYWORDS
angiogenesis, glioma, gene signature, prognosis, tumor immune microenvironment

Introduction

Glioma is the most common malignant tumor of the central nervous system (CNS), accounting for approximately 15% of all brain tumors (Ostrom et al., 2019). By degree of malignancy, gliomas are classified into low-grade gliomas (LGGs) and glioblastoma multiforme (GBM) (Louis et al., 2016). Despite the availability of a variety of treatment options including surgery, radiotherapy and chemotherapy, immunotherapy, and targeted therapy (Aldape et al., 2019), prognosis in glioma has remained poor; this is especially true in GBM patients, whose median survival time is < 15 months (Chen et al., 2017; Xu et al., 2020a). This poor prognosis is largely attributed to aberrant angiogenesis, high invasiveness, and therapeutic resistance (Furnari et al., 2007; Tan et al., 2018). According to previous research, gliomas with *IDH* mutation and 1p/19q codeletion have a relatively favorable prognosis (Eckel-Passow et al., 2015). The methylation status of the MGMT promoter has emerged as a key predictive biomarker of glioma and a potential predictor of response to temozolomide (Wick et al., 2014; Butler et al., 2020). However, additional research is needed to explore novel prognostic biomarkers and identify new therapeutic targets.

Angiogenesis refers to the formation of new blood vessels in the existing vasculature, which plays a pivotal role in many physiological and pathological processes such as embryonic development, wound healing, and tumor progression (Carmeliet, 2005). The pathophysiological processes of angiogenesis are reported to play critical roles in glioma development and therapeutic resistance (Onishi et al., 2011). Due to the important role of angiogenesis in gliomas, the use of angiogenesis-related genes (ARGs) to effectively stratify risk determining potential targets for individualized treatment is a promising research strategy. However, there have been few studies on the link between ARGs and prognosis in patients with glioma.

More recently, numerous studies have shown that the tumor immune microenvironment (TIME) plays a critical role in tumor progression and response to therapeutics (Quail and Joyce, 2017). Tumor-infiltrating immune cells can regulate tumor growth and invasion and are key components of the tumor microenvironment (TME) (Xu et al., 2020a; Xu Y. et al., 2020c). `The existing body of research on the TME suggests that immunotherapy is a promising method for the treatment of malignant tumors (Kruger et al., 2019; Xu et al., 2020b). In addition, the components of the TIME are closely correlated with the efficacy of immunotherapy.

In this study, we used data from the Cancer Genome Atlas (TCGA) and the Chinese Glioma Genome Atlas (CGGA) databases to explore the expression profiles and prognostic value of ARGs in gliomas. Then, based on ARG expression, we constructed clustering subgroups and risk models to verify the predictive value of ARGs in risk stratification and clinical outcome. We also evaluated the associations between the ARG expression risk signature and the immune microenvironment, tumor mutational burden (TMB), and immunotherapy response. Finally, to validate the clinical application of the ARG expression signature, a nomogram model was developed to predict the overall survival (OS) rates of glioma patients. The flow chart of this study is shown in Figure 1.

Materials and methods

Data resources

The TCGA dataset provided raw counts of RNA-sequencing data (FPKM values) and accompanying clinical information for glioma samples. The expression data and clinical information of the validation RNA-seq cohort CGGA693 were acquired from the CGGA website. We transformed the FPKM values into transcript per million (TPM) values (Wagner et al., 2012); all values of the expression data were $\log_2(x+1)$-transformed. The characteristics of patients in the TCGA and CGGA cohorts are summarized in Supplementary Table S1.

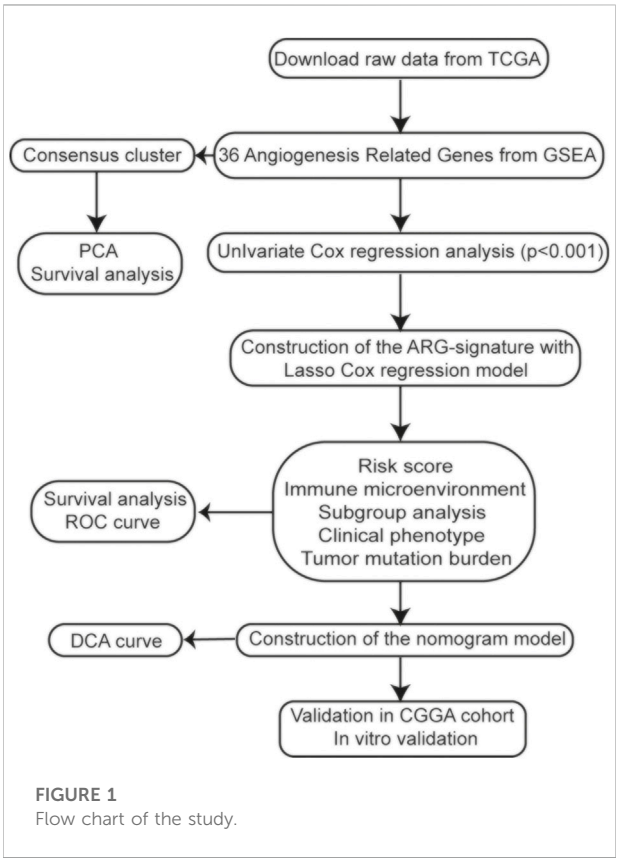


FIGURE 1
Flow chart of the study.

Consensus clustering analysis

We used the R package ConsensusClusterPlus, version 1.54.0, for consistency analysis. The maximum number of clusters was 6, and 80% of the total sample was drawn 100 times, clusterAlg = "hc," innerLinkage = 'ward.D2.' CDF and consensus matrices were used to calculate the appropriate number of subtypes. Then, we used PCA to detect differential gene expression between the two subtypes.

Construction of the angiogenesis-related gene signature

Univariate Cox regression analysis was performed to screen out ARGs significantly correlated with survival ($p < 0.001$). Next, biomarkers of the 28 ARGs were identified from the LASSO Cox regression algorithm using the glmnet package in R. We calculated the risk score of each glioma patient by the following formula:

$$Riskscore = \sum_{i=1}^{n} (Coef_i * x_i),$$

where $Coef_i$ is the coefficient of each ARG and x_i is the expression level of each ARG. In the risk score model, samples were subdivided into high- and low-risk groups according to the median risk score value.

Tumor-infiltrating immune microenvironment analysis

CIBERSORT is a deconvolution method for expression matrices of immune cell subsets (Newman et al., 2019). LM22 is a gene signature matrix that specifies the content of immune cell types. We used the CIBERSORT package in R to calculate the number of immune cells per sample, setting the permutation to 1,000 and selecting $p < 0.05$ as the screening threshold. The ESTIMATE algorithm was used to evaluate immune score, tumor purity, and stromal score (Yoshihara et al., 2013). We calculated abundances of immune infiltrates, including B cells, $CD4^+$ T cells, $CD8^+$ T cells, neutrophils, macrophages, and dendritic cells (DCs), using Tumor IMmune Estimation Resource (TIMER) (Li et al., 2017).

Single-sample gene set enrichment analysis

We used the ssGSEA method with the Gene Set Variation Analysis (GSVA) package in R to evaluate infiltration levels of different immune cells, the related expression pathways, and the activity of immune-related functions.

Tumor mutational burden analysis

We used the Maftools package to analyze and visualize somatic-mutation data in order to study the mutational landscapes of glioma patients (Mayakonda et al., 2018). TMB was defined as the total number of somatic mutations per million bases.

Survival analysis

We conducted Kaplan–Meier (KM) analysis to characterize the differences in survival of glioma patients using the R packages survival and survminer. The significance of differences in survival time was determined by using the log-rank test ($p < 0.05$).

Building and verification of the nomogram

The nomogram was constructed using the rms package in R. We created a calibration curve to examine the consistency between the actual survival rate and expected survival rate. We built the nomogram model based on our multivariate Cox regression results. We created calibration plots of the nomogram for 1-, 3-, and 5-year OS using the "calibrate" function in rms. Decision curve analysis (DCA) was used to assess the clinical net benefit.

Protein–protein interaction

The protein–protein interaction (PPI) analysis of ARGs was performed by using the STRING website (https://www.string-db.org/). The interaction analysis was conducted by Cytoscape software. The hub nodes were identified by the MCC method of cytoHubba plugin.

Cell culture

We cultivated the glioma cell lines U87 and LN229 in high-glucose Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS), 100 U/ml penicillin, and 100 μg/ml streptomycin at 37°C with 5% CO_2. *SPP1* small-interfering RNA (siRNA) sequences were as follows: si-*SPP1*-1: CCAGTTAAA CAGGCTGATT; si-*SPP1*-2: GTCTCACCATTCTGATGAA.

Western blotting

Western blot (WB) analysis was performed as previously reported (Han et al., 2017). Briefly, we extracted total proteins using a Total Cell Protein Extraction Kit (KeyGen Biotechnology, Nanjing, China). Equal amounts of protein were electrophoresed, transferred onto nitrocellulose membranes, and blocked with 2% bovine serum albumin (BSA). We used primary antibodies against *SPP1* (1:1,000; ab69498; Abcam, Cambridge, United Kingdom) to detect the expression of this protein. After washing them four times with Tris-buffered saline + Polysorbate 20 (TBST)/0.1% Tween-20, we incubated the membranes with the corresponding secondary antibody. A glyceraldehyde 3-phosphate dehydrogenase (GAPDH) protein band was used as a control to normalize protein levels. We visualized protein bands using a chemiluminescence kit (Beyotime Biotechnology, Beijing, China).

Cell viability assay

We inoculated the treated U87 and LN229 cells in 96-well plates at a density of 1×10^3 cells/well for 24, 48, 72, 96, and 120 h. The plates were examined using a cell viability assay kit (Promega Corp., Fitchburg, WI, United States) in accordance with the manufacturer's protocol, as described previously (Wang et al., 2021).

5-ethynyl-2′-deoxyuridine cell proliferation assay

We performed an EdU assay to visualize the proliferating cells and used a Click-iT EdU Alexa Fluor 488 Imaging kit (Invitrogen Corp., Carlsbad, CA, United States) to detect cell proliferation as per the manufacturer's instructions. We photographed EdU^+ cells under a fluorescence microscope and counted them using ImageJ software (National Institutes of Health [NIH], Bethesda, MD, United States).

Transwell invasion assay

We performed a transwell invasion assay according to previously described methods (Han et al., 2015). U87 and LN229 cell invasion was assessed using a Matrigel-coated filter over the lower compartment for 20 h. We counted the invading cells under a microscope (Olympus, Tokyo, Japan).

Co-culture

Glioma cells and human brain microvascular endothelial cells (hBMECs) were co-cultured in Boyden chambers. Briefly, hBMECs were cultured in 6-well plates, while glioma cells were seeded in chambers.

Tube formation assay

A pre-cooled 96-well plate was coated with 50 µl Matrigel (BD Biosciences, United States) per well and incubated at 37°C for 30 min. PBS was used to wash the tumor cells, and 0.25% trypsin was used for digestion. Cells were collected and counted using a hemocytometer after centrifugation. Then, the cells were resuspended with serum-free DMEM, and 2×10^4 cells/well were inoculated on the surface of Matrigel. After 12 h, tube formation was photographed using a microscope (Olympus, Tokyo, Japan). ImageJ software was used to quantify and analyze tubule intersections.

Statistical analysis

Statistical analyses and visualization were carried out in R. We performed time-dependent receiver operating characteristic (ROC) curve analysis to evaluate the predictive value of the constructed risk model using the R package survivalROC. The Wilcoxon test was used for comparisons between two groups, and the Kruskal–Wallis test was used for comparisons between multiple groups. A two-sided $p < 0.05$ was considered to be statistically significant.

Results

Consensus cluster analysis for angiogenesis-related gene expression profiles

The set of ARGs we obtained from Gene Set Enrichment Analysis—Hallmark, Angiogenesis (GSEA) included 36 genes that are upregulated in tumorigenic angiogenesis (Subramanian et al., 2005; Ren et al., 2020). We performed consensus clustering in the glioma patient training cohort to analyze the prognostic implications of the ARGs (Figure 2A). The empirical cumulative-distribution function (CDF) plot revealed the lowest rangeability at 0.2–0.8, with $k = 2$ (Figure 2A); the delta area scores were the highest also at $k = 2$ (Figure 2A). In addition, the maximum consistency was found at $k = 2$ in the consensus matrix plot (Figure 2A; Supplementary Figure S1). Therefore, $k = 2$ was shown to have the best clustering stability. Cluster 1 ($n = 260$) and cluster 2 ($n = 403$) were generated from a total of 663 patients. We used principal component analysis (PCA) to display differences in gene expression levels between the two subgroups (Figure 2A). The heatmap shows the expression pattern of 36 ARGs in clusters 1 and 2 (Figure 2B). We found

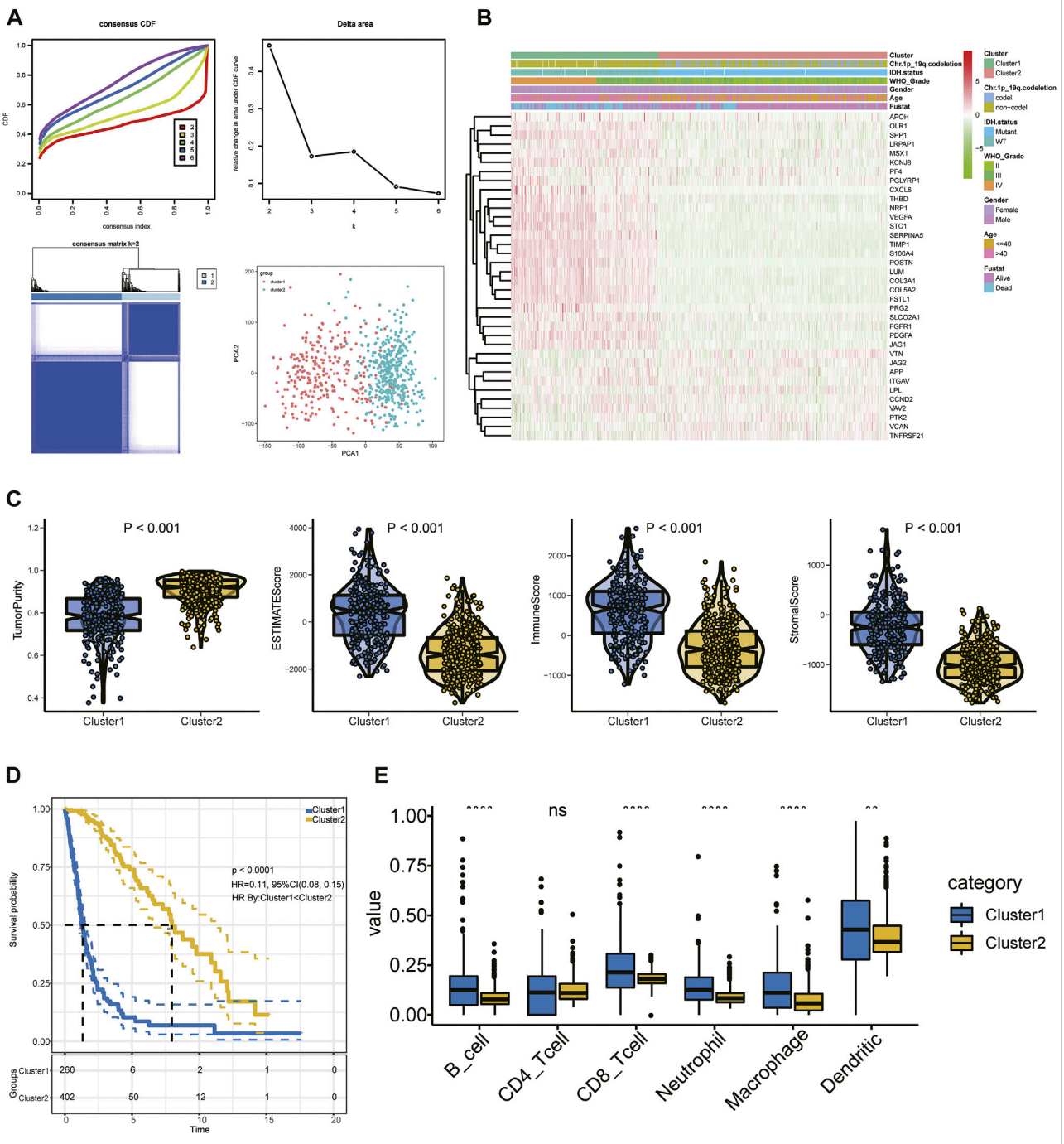


FIGURE 2
(A) Consensus clustering, CDF, and relative change in the CDF AUC with k = 2–6. **(B)** Heatmap of clinical information of the two clusters among 36 ARGs. **(C)** Tumor purity and ESTIMATE, stromal, and immune scores. **(D)** KM curve of glioma patients. **(E)** Content of six immune cells.

that immune score was significantly higher ($p < 0.05$), while tumor purity was significantly lower ($p < 0.05$) in cluster 1 than in cluster 2 (Figure 2C). Furthermore, a KM curve showed that the OS outcome of cluster 1 was worse than that of cluster 2 (Figure 2D). In addition, cluster 1 had significantly higher abundances of B cells, $CD8^+$ T cells, neutrophils, macrophages, and DCs than cluster 2 ($p < 0.05$), while there was no between-cluster difference in $CD4^+$ T cells (Figure 2E). These results indicated that the cluster assignment based on ARGs was closely related to prognosis and TIME in glioma.

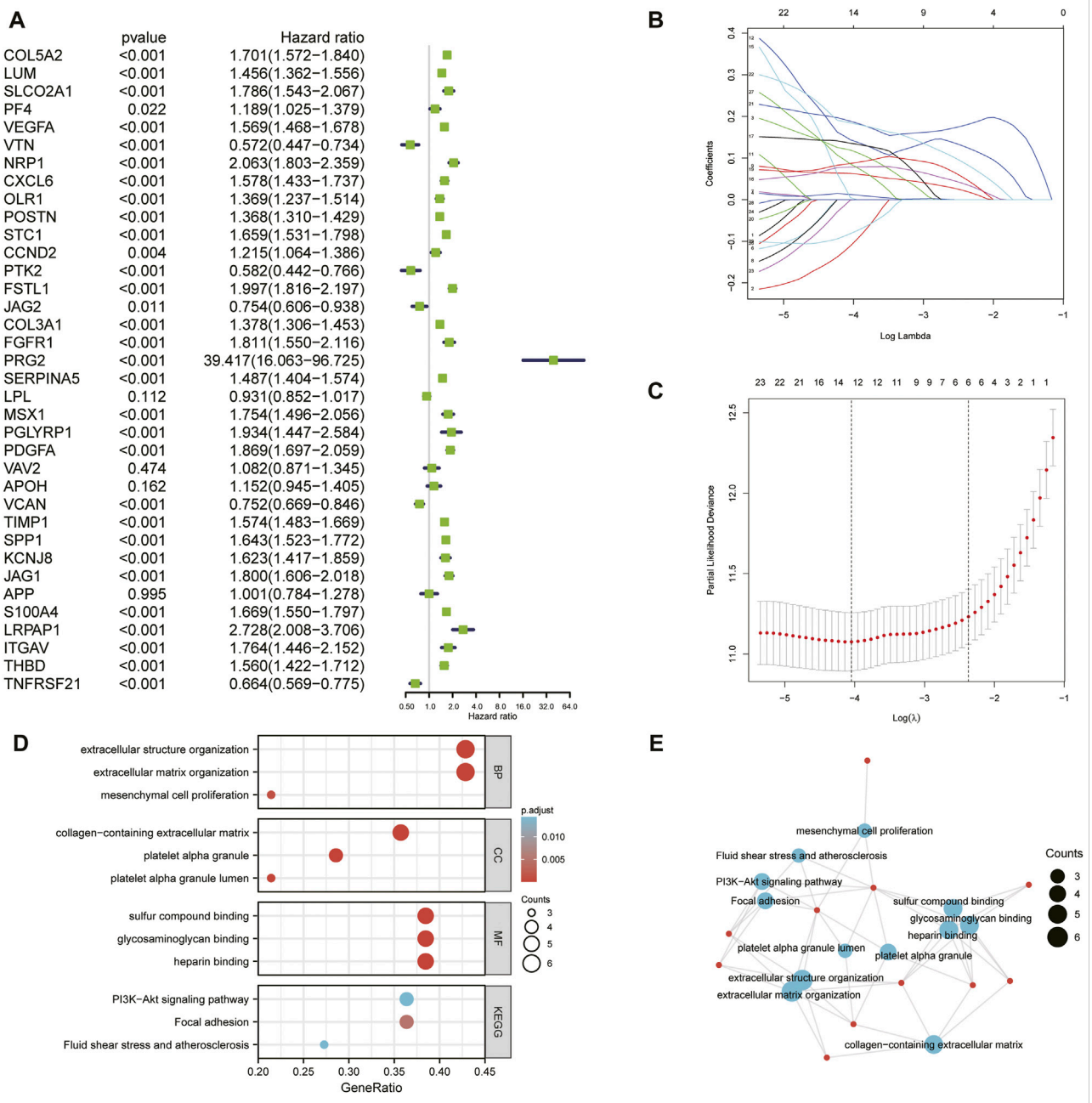


FIGURE 3
(A) Univariate Cox regression analysis of the 36 ARGs in the TCGA cohort. **(B)** LASSO coefficient profiles of the common genes. **(C)** Cross-validation for tuning parameter screening in the LASSO regression model. **(D)** GO and KEGG enrichment analysis across the 14 genes. **(E)** Functional-enrichment map of pathways of the 14 ARGs.

Establishment and validation of the risk signature based on angiogenesis-related gene expression

First, we conducted univariate Cox regression analysis to screen out 29 OS-related ARGs ($p < 0.001$) in the TCGA cohort (Figure 3A). Subsequently, we selected these genes to conduct an additional least absolute shrinkage and selection operator (LASSO) Cox regression analysis (Figures 3B,C). The formula was as follows: risk score = (*LUM* × −0.11114) + (*SLCO2A1* × 0.11913) + (*VEGFA* × 0.01235) + (*POSTN* × 0.06287) + (*FSTL1* × 0.14389) + (*PRG2* × 0.00485) + (*SERPINA5* × 0.07829) + (*MSX1* × 0.13564) + (*PDGFA* × 0.08695) + (*TIMP1* × 0.1885) + (*SPP1* × 0.18423) + (*KCNJ8* × −0.00092) + (*ITGAV* × 0.08581)

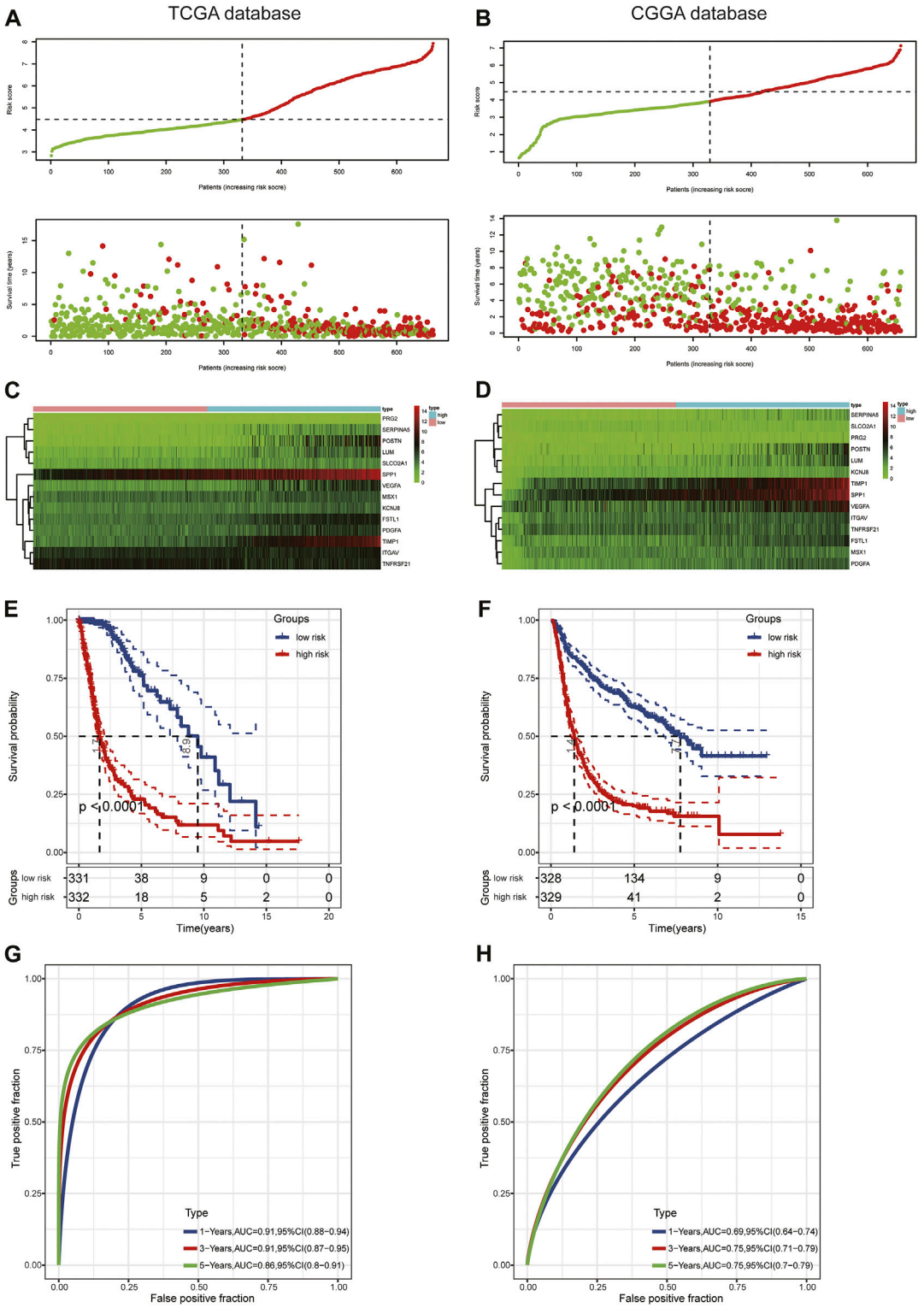


FIGURE 4
Prognostic value of the risk score in TCGA and CGGA. **(A,B)** Distribution of risk score and survival status. **(C,D)** Expression pattern of 14 ARGs in the high- and low-risk groups. **(E,F)** KM analysis of the risk model. **(G,H)** Time-dependent ROC curve analysis of the risk model.

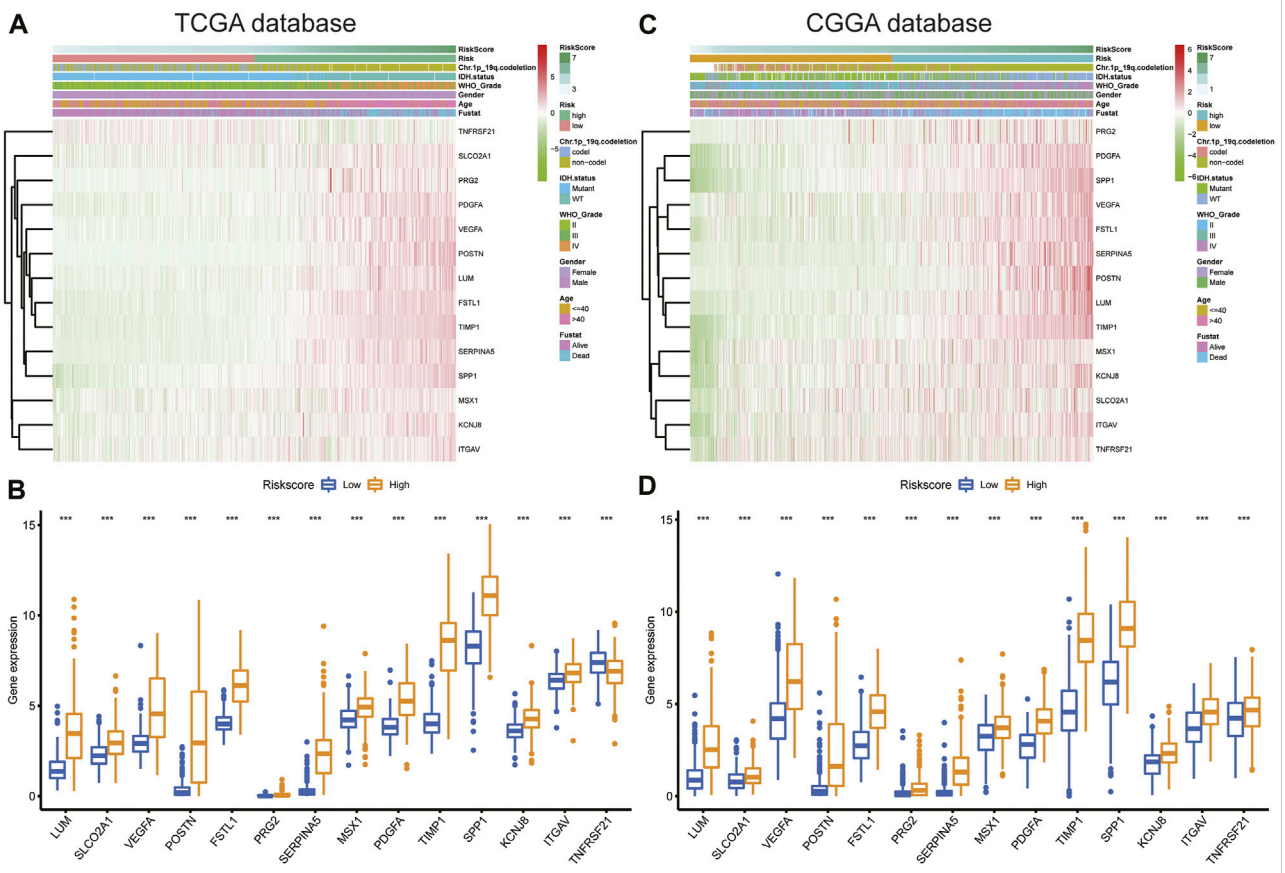


FIGURE 5
(A,C) Heatmap of the 14-ARG expression pattern in clinicopathologic characteristics and risk score in the TCGA and CGGA databases. **(B,D)** Expression differences in the 14 ARGs between the low- and high-risk groups in the TCGA and CGGA databases.

+ (*TNFRSF21* × −0.0817). GO and KEGG enrichment analysis was performed by R package "clusterProfiler" (Yu et al., 2012). These genes were shown to be involved in extracellular structure organization and the PI3K-Akt signaling pathway (Figures 3D,E). Differential analysis was performed to detect 14 ARGs (Supplementary Figure S2). Patients in the training cohort (TCGA) were divided into high- and low-risk groups based on the median risk score. According to our findings, the number of patients who died increased as their risk score increased (Figures 4A,B). Differential expression levels of the 14 ARGs in the high- and low-risk groups are shown in heatmaps (Figures 4C,D). To evaluate the role of the 14-ARG signature in glioma, we drew KM curves for the high- and low-risk groups of the TCGA cohort (Figure 4E). These two subgroups significantly differed in OS ($p < 0.0001$). Thereafter, we used a time-dependent ROC curve to predict the efficacy of the risk signature. The area under the curve (AUC) of the prediction model was 0.91 over 1 year, 0.91 over 3 years, and 0.86 over 5 years in the TCGA training cohort (Figure 4G).

To assess the predictive value of the risk model, we used the risk score algorithm in the CGGA cohort. The results in the validation cohort revealed that glioma patients in the high-risk group had worse survival rates than those in the low-risk group (Figure 4F). The AUCs for 1-, 3-, and 5-year survival were 0.69, 0.75, and 0.75, respectively (Figure 4H). These findings suggested that the 14-ARG risk model could accurately predict the prognoses of patients with glioma.

Association between angiogenesis-related gene risk signature and clinical information

Expression of the 14 ARGs in low- and high-risk patients in the TCGA and CGGA datasets is depicted by heatmaps (Figures 5A,C). Other than those of *TNFRSF21*, expression of the 13 other ARGs increased significantly ($p < 0.05$) in the high-risk group (Figure 5B) of the TCGA cohort. All 14 ARGs were highly expressed in the high-risk group in the CGGA database ($p <$

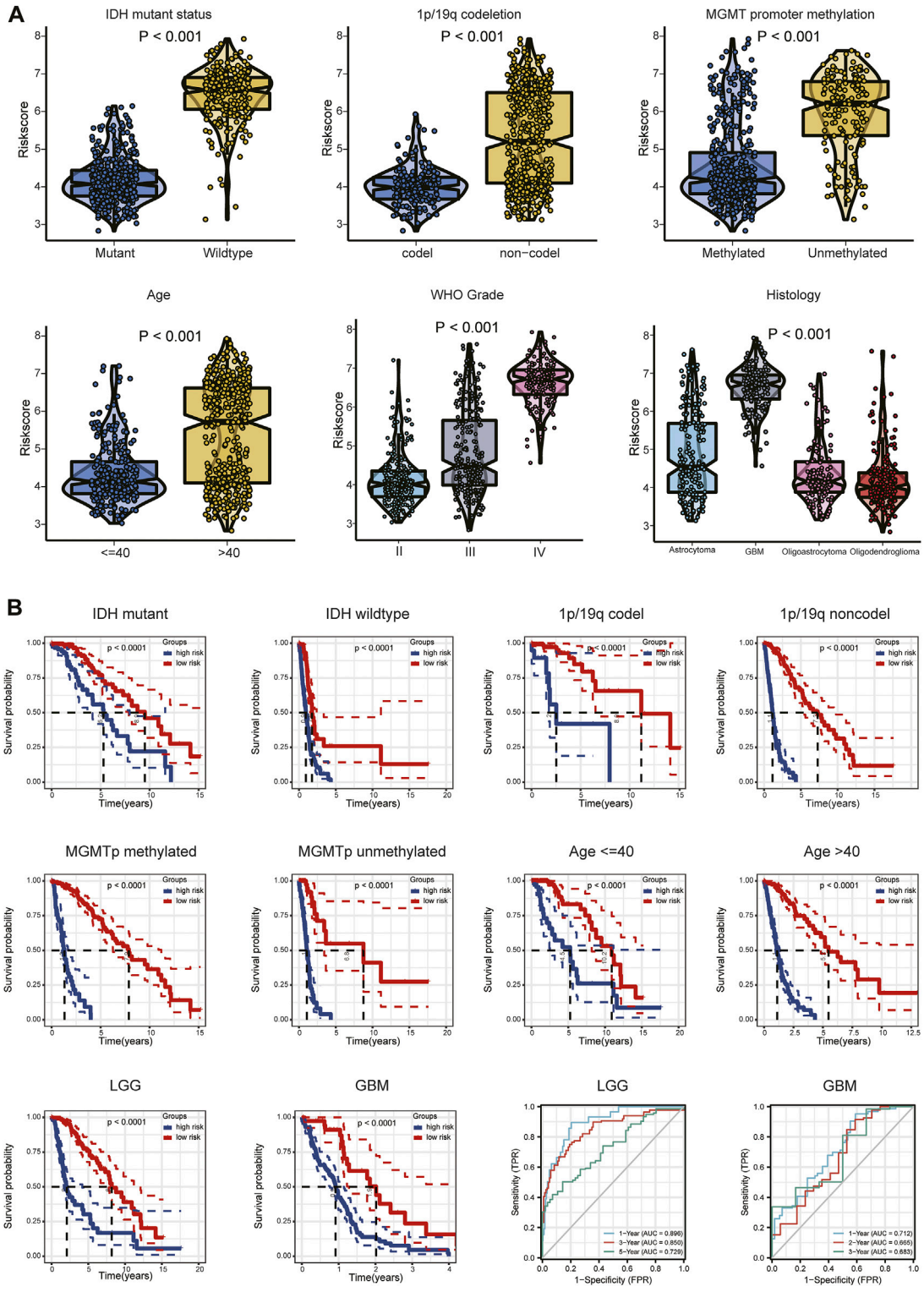


FIGURE 6
(A) Relationship between risk score and each clinicopathological characteristic (*IDH*-mutant status, 1p/19q codeletion, MGMT promoter methylation, age, WHO grade, and histology). **(B)** KM analyses of patients in the CGGA dataset stratified by *IDH*-mutant status, 1p/19q codeletion, MGMT promoter methylation, age, and WHO grade in the TCGA cohort. ROC curve analysis of the risk model in predicting 1-, 3- and 5-year OS in the TCGA–LGG cohort and 1-, 2- and 3-year OS in the TCGA–GBM cohort.

0.05; Figure 5D). We also performed survival analysis of single ARGs in glioma patients (Supplementary Figures S3,S4). The results showed that for glioma patients in the TCGA cohort, all of the ARGs were prognostic-risk factors, except for *TNFRSF21*. Thereafter, we evaluated differences in risk score between different clinicopathological characteristics of glioma patients in the training and validation cohorts, including *IDH* mutation status, 1p/19q codeletion, MGMT promoter methylation, age, WHO grade, and histology. The results showed that in the TCGA dataset, the risk score was elevated in the *IDH* wild-type (WT), 1p/19q non-codeletion subtype, MGMT promoter unmethylated subtype, older patients, and high-grade glioma ($p < 0.05$); we validated these results in the CGGA dataset (Figure 6A; Supplementary Figure S5A). Next, we drew KM curves of the risk score signature stratified by *IDH*-mutant status, 1p/19q codeletion, MGMT promoter methylation, age, and WHO grade in the glioma patients of the training and validation cohorts. The KM curve suggested the predictive value of the ARG risk score signature in prognosis in the LGG and GBM subgroups (Figure 6B; Supplementary Figure S5B). The results demonstrated the power of the ARG risk score signature's prognostic value in the glioma subgroups of the TCGA cohort (Figure 6B), and these results were consistent in the CGGA cohort (Supplementary Figure S5B).

Because different grades of glioma have different clinical features and prognoses, we performed subgroup analyses of LGG and GBM. The relationships between risk score and each clinical characteristic (*IDH*-mutant status, 1p/19q codeletion, MGMT promoter methylation, age) in the TCGA/CGGA-LGG and TCGA/CGGA-GBM subgroups are shown in Supplementary Figures S6A,S6D and in Supplementary Figures S7A,S7D, respectively. Tumor purity was significantly higher ($p < 0.05$) and ESTIMATE, immune, and stromal scores significantly lower ($p < 0.05$) in the low-risk group in the LGG and GBM subgroups (Supplementary Figures S6B,S6E, Supplementary Figures S7B,S7E). Expression differences of the 14 ARGs between the high- and low-risk groups of the LGG and GBM subgroups are shown in Supplementary Figures S6C,S6F and in Supplementary Figures S7C,S7F. The ROC curve showed the efficiency of the risk signature in these two subgroups. The AUC of the prediction model was 0.896 over 1 year, 0.850 over 3 years, and 0.729 over 5 years in the LGG subgroup and 0.712 over 1 year, 0.665 over 2 years, and 0.683 over 3 years in the GBM subgroup (Figure 6B; Supplementary Figure S5B). These results indicated the predictive stability of the 14-ARG risk score model's prognostic value in both these subgroups.

Next, we performed univariate and multivariate Cox regression analyses in the TCGA and CGGA cohorts to assess the independent prognostic value of the ARG risk signature. We observed that in univariate analysis, age, WHO grade, *IDH* status, chromosome 1p/19q status, and risk score were significantly correlated with prognosis in both the TCGA and CGGA cohorts (Figures 7A,C). However, multivariate analysis indicated that age, grade, and risk score were independent prognostic factors in the TCGA cohort (Figure 7B; $p < 0.05$). In the validation cohort (CGGA), we also found that risk score was an independent prognostic factor (Figure 7D; $p < 0.05$).

Furthermore, we compared the prognostic predictive abilities of 20 different risk signatures of gliomas in TCGA from published articles, including inflammatory response-related gene (IRRG) signature (Yan et al., 2022), DNA damage and repair-related gene (DDRRG) signature (Li et al., 2022c), CXCR members signature (He et al., 2022), pyroptosis-related gene signature (Zhang M. et al., 2021b; Chao et al., 2022; Yang et al., 2022; Zhang et al., 2022), ECM-related gene (ECMRG) signature (Li et al., 2022b), tripartite motif (TRIM) family gene signature (Xiao et al., 2022), antigen presentation machinery (APM) signature (Chen et al., 2022), natural killer cell-related gene (NKRG) signature (Li C. et al., 2022a), IL-4-related gene (IL4RG) signature (Qi et al., 2022), hypoxia-related gene (HRG) signature (Gao et al., 2021), S100 family-based signature (Hu et al., 2021), TIME signature (Zhang C. et al., 2021a), focal adhesion-related gene (FARG) signature (Li et al., 2021), m6A RNA methylation regulator signature (Cong et al., 2021), HDAC1-related signature (Fan et al., 2021), RNA-binding protein (RBP)-based signature (Chen et al., 2021a) and ferroptosis-related gene (FRG) signature (Chen et al., 2021b). The results of univariate and multivariate Cox analyses showed that our ARG signature had independent predictive ability ($p < 0.001$, Table 1).

Based on the abovementioned comprehensive analyses, we considered the effect of risk score on prognosis to be accurate and stable.

Angiogenesis-related gene risk signature and the tumor immune microenvironment

The heatmap of immune responses based on the ESTIMATE algorithms and single-sample GSEA (ssGSEA) is depicted in Figure 8A. Tumor purity was substantially lower ($p < 0.05$) in the high-risk group, but ESTIMATE, immune and stromal scores were significantly higher (Figure 8B). We calculated the proportions of 22 types of immune cells in each glioma sample based on the CIBERSORT algorithm. Next, we compared differences in proportions of immune cells between the high- and low-risk groups in the TCGA database. Abundances of $CD8^+$ T cells, follicular helper T (T_fh) cells, regulatory T cells (Tregs), gamma delta (γδ) T cells, resting natural-killer (NK) cells, M0, M1, and M2 macrophages, and neutrophils were significantly more enriched in the high-risk than in the low-risk group (Figure 8C). Additionally, we identified two immune subtypes based on immune-genomic profiling of 29 immune signatures in ssGSEA. We found a significantly higher risk score in the immunity-high subtype than the immunity-low subtype (Figure 8D). We also compared six immune cell types *via* the TIMER algorithm,

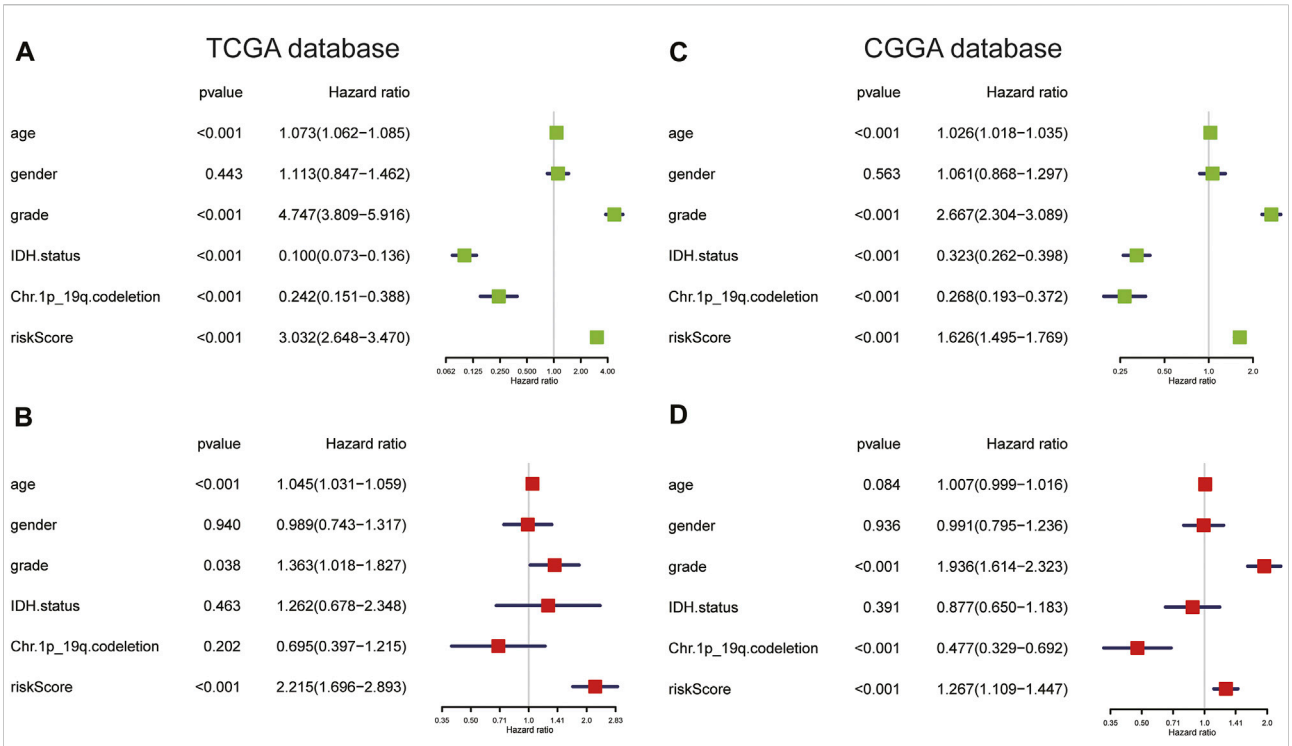


FIGURE 7
(A,B) Univariate and multivariate Cox regression analyses in the TCGA cohort. **(C,D)** Univariate and multivariate Cox regression analyses of clinicopathologic features in the CGGA cohort.

and results showed that abundances of B cells, $CD8^+$ T cells, neutrophils, macrophages, and DCs were significantly higher in the high-risk group (Figure 8E). We obtained similar TIME infiltration results in the validation cohort (Supplementary Figure S8), indicating greater infiltration of $CD8^+$ T cells, T_fh cells, Tregs, and M0 macrophages in the high-risk group (Supplementary Figure S8C), and risk score remained higher in the immunity-high subtype (Supplementary Figure S8D). These results demonstrated that the ARG risk signature was closely associated with infiltration of immune cells.

Angiogenesis-related gene risk signature and mutational profile

The mutational landscapes between the low- and high-risk groups of each glioma patient in TCGA were analyzed and are displayed as a waterfall plot (Figures 9A,B). Compared with the low-risk group, TMB was significantly high ($p < 0.001$) in the high-risk group (Figure 9C). A log rank test and the KM curve showed that the high-TMB group had worse survival outcomes than the low-TMB group ($p < 0.001$; Figure 9D). We also drew the survival curve of the TMB combined risk score (Figure 9E); the results showed that the high-TMB plus high-risk score group had a worse survival outcome ($p < 0.001$).

Angiogenesis-related gene risk signature and immunotherapy

The association between risk score and immunotherapeutic effect was also explored. We found that risk scores were positively correlated with expression of crucial immune checkpoints (*B7H3*, *PD-L1*, *PD-L2*, *HAVCR2*, *LAG-3*, *PD-1*, *CTLA4*, and the inflammatory factors *HLA-A*, *HLA-B*, and *HLA-C*) in the TCGA and CGGA databases (Figures 10A,B). Furthermore, we evaluated immune checkpoint and *HLA* complex expression levels. The high-risk group of the training and validation cohorts had considerably greater expressions of both. ($p < 0.05$; Figures 10C,D). Collectively, the results suggested that risk stratification could help predict the effect of immunotherapy in gliomas.

Construction and validation of the prognostic-nomogram model

To evaluate the prognostic significance of the ARG signature in glioma patients, we established a nomogram model based on age, WHO grade, and risk score (Figure 11A; Supplementary Figure S9A) using our multivariate-analysis results. The C-index of the nomogram model was generated to assess discriminating

TABLE 1 Univariate and multivariate Cox regression analyses of different risk signatures.

Characteristics	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	*p* Value	Hazard ratio (95% CI)	*p* Value
Our ARG signature	3.032 (2.648–3.470)	< **0.001**	3.019 (1.808–5.041)	**<0.001**
IRRG signature	13.574 (9.597–19.200)	< **0.001**	0.785 (0.342–1.804)	0.569
DDRRG signature	6.885 (5.419–8.748)	< **0.001**	2.899 (1.363–6.165)	**0.006**
CXCR member signature	1.251 (1.089–1.438)	**0.002**	0.971 (0.838–1.125)	0.698
PRG signature (Chao B *et al.*)	6.134 (4.858–7.745)	< **0.001**	0.714 (0.387–1.321)	0.283
PRG signature (Yang Z *et al.*)	2.555 (2.257–2.892)	< **0.001**	0.971 (0.698–1.350)	0.860
PRG signature (Zhang M *et al.*)	2.218 (2.013–2.444)	< **0.001**	1.165 (0.825–1.645)	0.385
PRG signature (Zhang Y *et al.*)	2.751 (2.413–3.135)	< **0.001**	1.177 (0.828–1.675)	0.364
ECMRG signature	5.518 (4.477–6.800)	< **0.001**	0.455 (0.182–1.136)	0.092
TRIM family gene signature	23.500 (14.501–38.083)	< **0.001**	0.904 (0.351–2.332)	0.835
APM signature	4.157 (3.333–5.185)	**<0.001**	0.578 (0.341–0.979)	**0.041**
NKRG signature	1195154.632 (130202.538–10970558.754)	< **0.001**	5.857 (0.187–183.904)	0.315
IL4RG signature	266.447 (124.392–570.730)	< **0.001**	0.738 (0.117–4.646)	0.746
HRG signature	2.974 (2.532–3.495)	< **0.001**	0.923 (0.681–1.251)	0.606
S100 family-based signature	2.833 (2.475–3.244)	< **0.001**	0.784 (0.524–1.172)	0.235
TIME signature	5.365 (4.355–6.607)	< **0.001**	1.365 (0.704–2.646)	0.357
FARG signature	2.974 (2.502–3.535)	< **0.001**	0.689 (0.510–0.931)	**0.015**
m6A RNA methylation regulators signature	3.852 (3.236–4.586)	< **0.001**	0.845 (0.562–1.269)	0.416
HDAC1-related signature	3.605 (3.033–4.284)	< **0.001**	1.158 (0.796–1.685)	0.444
RBP-based signature	3.130 (2.673–3.664)	< **0.001**	1.081 (0.808–1.445)	0.602
FRG signature	2.786 (2.456–3.159)	< **0.001**	1.569 (1.053–2.338)	**0.027**

The bold values are $p < 0.05$.

abilities, and it performed well (TCGA training cohort, 0.875; CGGA validation cohort, 0.735). In the TCGA and CGGA cohorts, the calibration curves revealed a favorable consistency between expected and observed survival rates (Figures 11B–D; Supplementary Figures S9B–D). In addition, we used DCA to examine the suitability of the nomogram in clinical settings. The model exhibited an excellent net benefit (Figures 11E–G; Supplementary Figures S9E–G). Taken together, the results described above suggested that the nomogram model had good reliability in predicting OS in glioma patients.

Knockdown of *SPP1* significantly inhibited cell proliferation, invasion, and angiogenesis

SPP1 was overexpressed in the high-risk group of glioma patients and was correlated with poor prognosis. The results of PPI analysis and the MCC method of cytoHubba suggested *SPP1* may be the hub gene (Figure 12A). In the U87 and LN229 glioma cell lines, we determined the role of *SPP1* using *in vitro* experiments. SiRNA was used to reduce expression of *SPP1* in both U87 and LN229 cells; *SPP1* protein expression levels are shown in Figure 12B. We used a cellular-viability assay to analyze the effects of *SPP1* on the proliferation of U87 and LN229 cells. The results, which were presented as the mean ± standard deviation (SD) of three independent experiments, suggested that *SPP1* knockdown significantly reduced the viability of glioma cells (Figure 12C; $p < 0.05$). Meanwhile, the results of EdU assay suggested that *SPP1* inhibited the proliferation capacity of the glioma cell lines (Figure 12D). Transwell experiments suggested that knockdown of *SPP1* could also inhibit migration and invasion of U87 and LN229 cells Figure 12E). hBMECs co-cultured with si-*SPP1* glioma cells showed attenuated network formation when compared with controls (Figure 13), which suggested knockdown of *SPP1* inhibited angiogenesis.

Discussion

Despite advances in surgical and medical treatment, glioma remains a fatal disease. Numerous studies indicate that aberrant angiogenesis is involved in the processes of tumorigenesis, development, invasion, and poor prognosis in glioma (Tan et al., 2018). To date, there are still few studies on ARG in

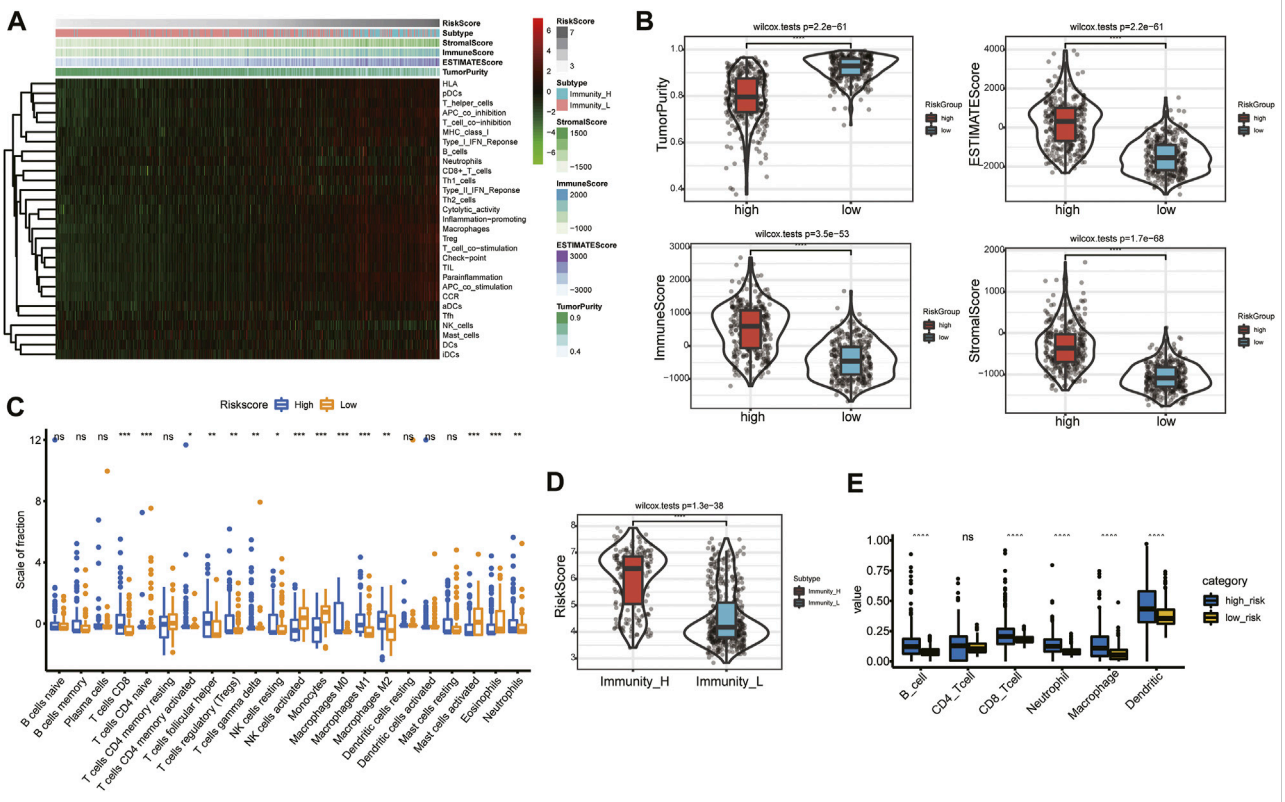


FIGURE 8
Relationship between risk signature and TIME in the TCGA database. **(A)** Heatmap of risk score and the two immunity subtypes based on ssGSEA. **(B)** Comparison of tumor purity and of ESTIMATE, immune, and stromal scores in the high- and low-risk groups. **(C)** Association between immune cells and the risk signature. **(D)** Comparison of risk score between the immunity-high and immunity-low subtypes. **(E)** Abundances of six immune cells in the high- and low-risk groups.

glioma (Biterge-Sut, 2020; Wang et al., 2022). Two major aspects of glioma biological processes that contribute to treatment resistance are abnormal formation of new blood vessels via angiogenesis and invasion of glioma cells along white-matter tracts (Carmeliet, 2005; Onishi et al., 2011). Although using immunohistochemistry (IHC) to analyze the expression level of a single angiogenesis gene is convenient (Tan et al., 2018; Peng et al., 2021), multi-gene signature analysis can reveal the complex interactions among various factors that affect angiogenesis in the pathophysiology of gliomas. Therefore, application of multi-gene methods might help researchers better describe the characteristics of tumor biology, thereby guiding clinical decision-making for accurate cancer diagnosis and treatment. The effectiveness of single-ARG targeted treatment is still limited (Onishi et al., 2011), suggesting that angiogenesis in glioma likely results from multiple genes and factors and that exploration of multi-gene signatures might provide guiding significance for multi-target combined therapy.

In this study, we performed consensus clustering based on the ARG expression level to create two clusters. KM analysis showed that glioma patients in cluster 1 had unfavorable clinical outcomes. Moreover, immune cell infiltration in cluster 1 was greater than that in cluster 2. These results indicated that high immune scores and high infiltration of immune cells were correlated with poor prognosis, which was consistent with that in previous studies (Deng et al., 2020; Tian et al., 2020; Xu et al., 2021). Next, we identified 14 ARGs of significance and applied them to build a risk model by combining LASSO and Cox regression analyses. The risk score showed a favorable predictive value for the survival rate of glioma patients in the training and validation cohorts. Moreover, the risk score was found to be an independent predictor of glioma prognosis in multivariate Cox regression analyses. Furthermore, we established and validated a nomogram model to predict OS in glioma. The calibration curve revealed high concordance between predicted and actual OS rates, indicating good prediction performance of the nomogram model.

The biological functions of 14 ARGs have been moderately studied in various cancers, but not as much in gliomas. Crocker et al., (2011) found that *TIMP-1* serum level is positively correlated with *TIMP-1* expression in tumor tissue and inversely correlated with survival time of glioma

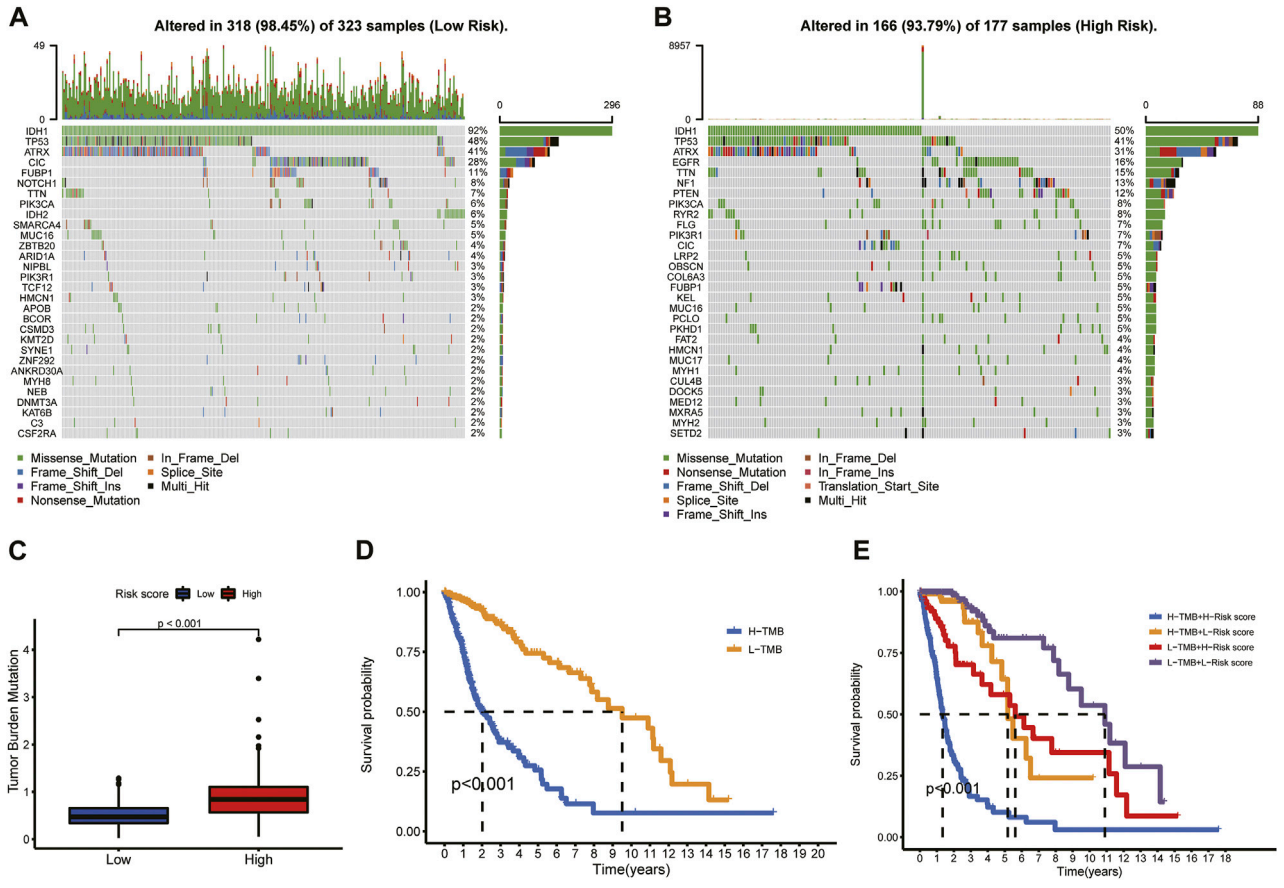


FIGURE 9
Mutational profile and TMB in the low- and high-risk groups. **(A)** Mutational profile in the low-risk group. **(B)** Mutational profile in the high-risk group. **(C)** Difference in TMB between low- and high-risk groups. **(D)** KM analysis of the high- and low-TMB groups. **(E)** Survival curve of the TMB combined risk score.

patients. *VEGFA* is a critical target of anti-angiogenic treatment for a variety of malignant tumors, including gliomas, since it is a fundamental mediator of tumor angiogenesis (Tamura et al., 2019). In addition to angiogenesis, *VEGFA* can inhibit the maturation of DCs to inhibit tumor immune response and induce immunosuppressive cells (Lindau et al., 2013). Previous research studies have shown that elevated *VEGFA* expression levels are related to poor prognosis in many tumors, including gliomas (Hicklin and Ellis, 2005). Reddy et al., (2008) found that overexpression of *FSTL1* is a biomarker of poor prognosis in GBM patients, and Jin et al., (2017) demonstrated that this gene is a critical modulator that promotes cell proliferation and cell cycle progression. Overexpression of *SPP1* is associated with poor OS in patients with glioma (Chen et al., 2019). The results of our functional experiments showed that *SPP1* knockout could inhibit the proliferation, invasion, and angiogenesis of glioma cell lines U87 and LN229. Therefore, we believe that *SPP1* might affect the prognosis of glioma by helping regulate angiogenesis and cell proliferation. The abovementioned evidence indicated that the 14 ARGs might play important roles in angiogenesis, invasiveness, and the TIME of gliomas. This also suggested that the ARG risk signature could help support clinical decision-making in glioma patients.

Previous studies have shown that immune infiltration plays an important role in determining therapeutic effect and prognosis in glioma patients (Gentles et al., 2015; Pereira et al., 2018; Kruger et al., 2019; Xu et al., 2020b). Tumor angiogenesis facilitated by hypoxia in the TIME leads to an antitumor immune response (Abou Khouzam et al., 2020). Macrophages are abundant cell components in the glioma microenvironment, which can promote proliferation, invasion, and migration of glioma (Uneda et al., 2021). Researchers have found that a high level of

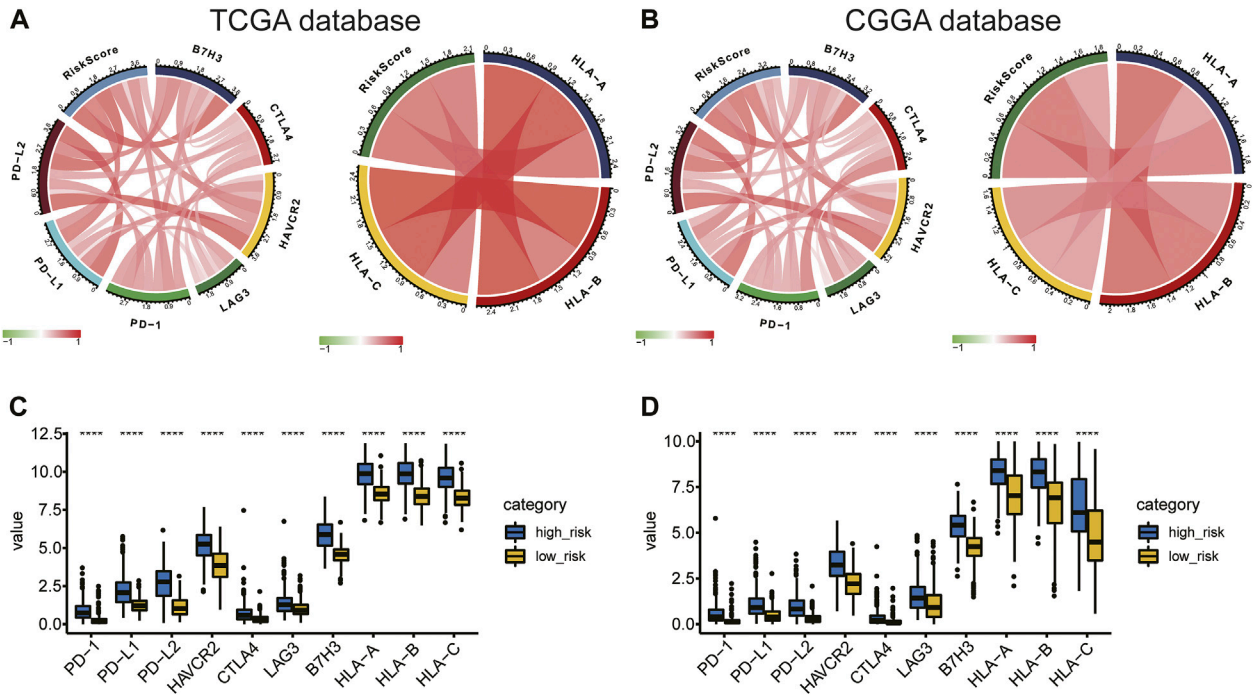


FIGURE 10
(A,B) Correlation of risk score to immune checkpoints and *HLA* complex expression levels. **(C,D)** Difference in expression of immune checkpoints and the *HLA* complex between the high- and low-risk groups.

infiltrating $CD8^+$ T cells is correlated with poor prognosis in glioma (Zhai et al., 2017; Weenink et al., 2019; Guo et al., 2020). Therefore, we further explored the relationship between immune cell infiltration and risk stratification. Data from the ESTIMATE algorithm showed that ARG risk stratification was negatively correlated with tumor purity and positively correlated with immune and stromal scores, which suggested higher infiltration levels of immune and stromal cells in the TME of the high-risk group. Numerous studies have shown that TAMs might promote the proliferation and progression of gliomas by enhancing immunosuppression, migration, invasion, and angiogenesis (Li and Graeber, 2012; Coniglio and Segall, 2013; Kennedy et al., 2013; Zhang Y. et al., 2021c). In our study, we found that the high-risk group had a higher infiltration of immunosuppressive cells such as M2 macrophages and Tregs, which create an immunosuppressive microenvironment and inhibit NK cell activation. The abundance of activated NK cells in the high-risk group was lower than that in the low-risk group. In general, we speculate that the poor prognosis of glioma patients in the high-risk group might be related to the tumor immunosuppressive microenvironment.

Multiple studies have reported that glioma acquires aggressive characteristics depending on a series of genome alterations (Kim et al., 2015; Yin et al., 2020). TMB has become a novel potential biomarker for predicting the efficacy of immune checkpoint therapy in many cancers (Braun et al., 2016; Chan et al., 2019). We explored the mutational profiles and TMBs of the high- and low-risk groups to investigate the predictive value of the risk model. We found that TMB increased significantly in the high-risk group and that patients with high TMB had poor prognoses. Consistent with our findings, Yin et al., (2020) found that TMB is negatively correlated with OS in glioma patients. Previous studies have suggested that immune checkpoints and the *HLA* complex have been implicated in the treatment response and prognosis of glioma (Luoto et al., 2018; Cloughesy et al., 2019; Feng et al., 2019). Kim et al., (2020) found that *HAVCR2* (*TIM-3*) plays specific intracellular and intercellular immunoregulatory roles in the TME of gliomas. Studies have shown that the *HLA* level is positively related with development of gliomas (Machulla et al., 2001). In this study, risk score was positively correlated with expression of immune checkpoint molecules and *HLA* complex. These findings demonstrated the 14-ARG risk model's accuracy in the prediction of the TIME of glioma, which therapeutic targets based on this signature might alter. The ARG expression signature could be used to predict clinical prognosis and efficacy of immunotherapy in glioma

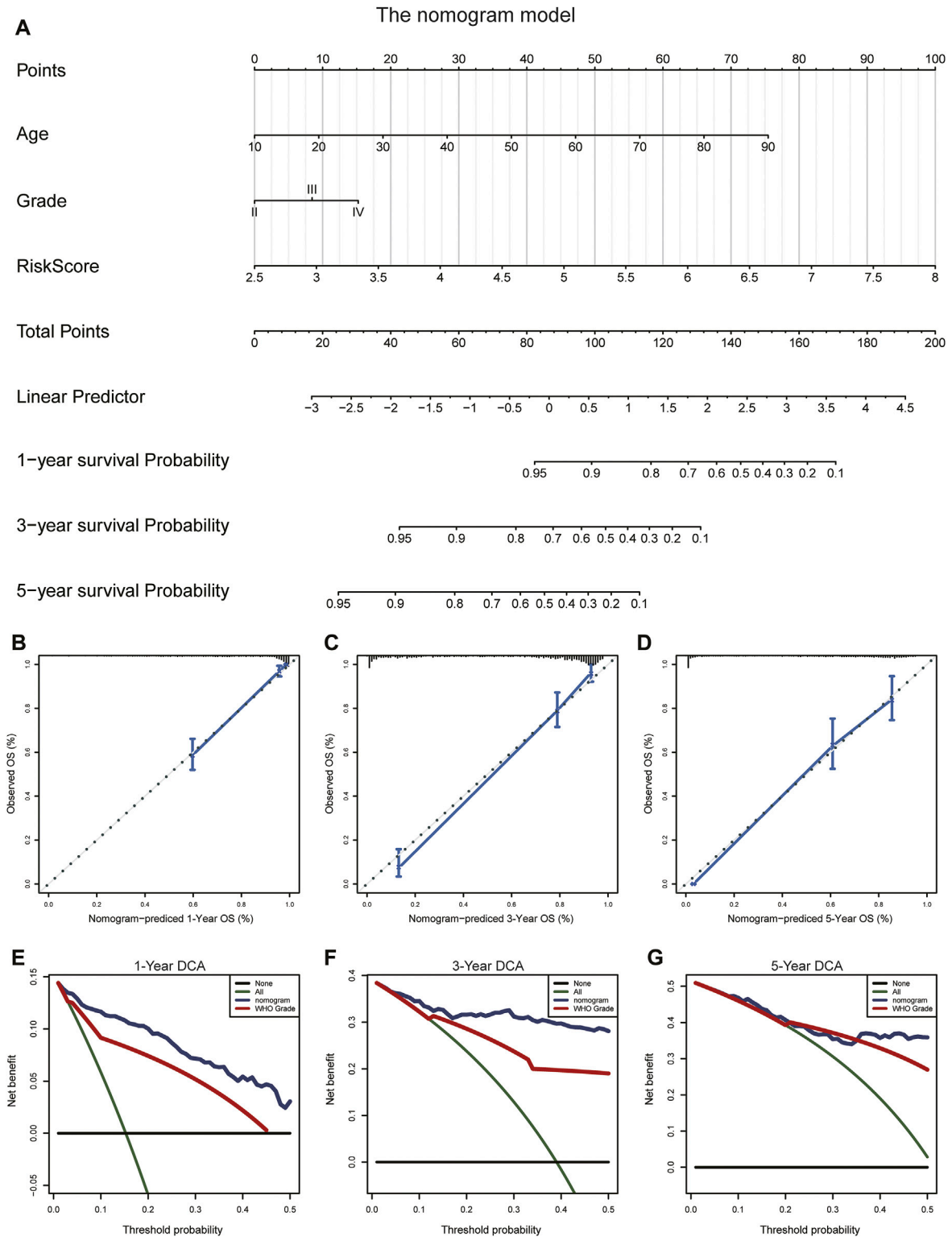


FIGURE 11
Construction and validation of the nomogram to predict OS in glioma patients. **(A)** The nomogram was established using age, WHO grade, and the ARG risk signature in the TCGA cohort. **(B–D)** Calibration curve of the nomogram for predicting the probability of OS at 1, 3, and 5 years in the TCGA cohort. **(E–G)** DCA of the OS-related nomogram at 1, 3, and 5 years in the TCGA cohort.

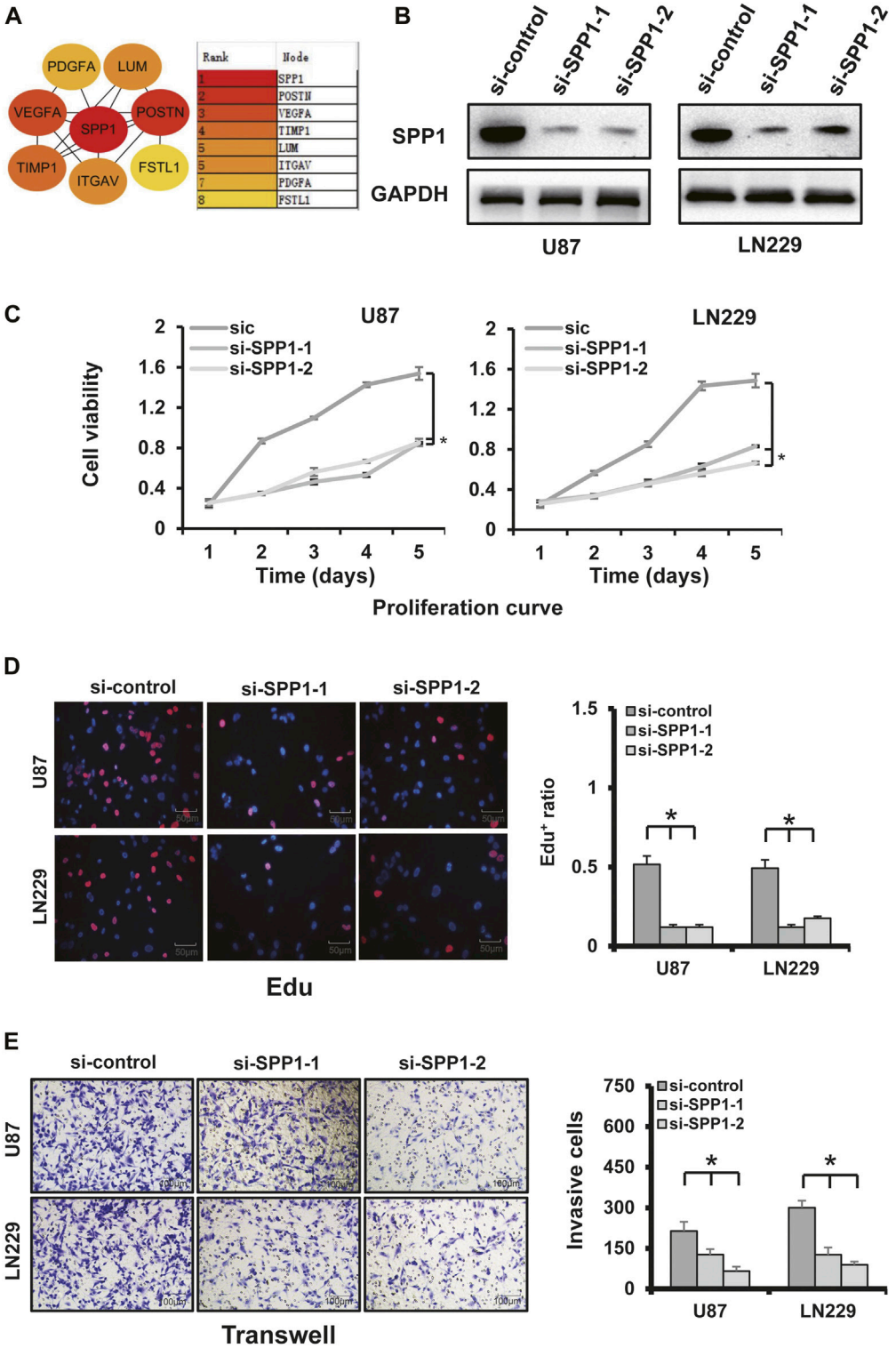


FIGURE 12
SPP1 experiments. **(A)** PPI analysis and the MCC method of cytoHubba showed that *SPP1* had the highest hub node score. **(B)** *SPP1* knockdown using two independent *SPP1* siRNAs (si-*SPP1*-1, si-*SPP1*-2) in U87 and LN229 cells was evidenced by WB analysis. *GAPDH* was using as loading control. **(C)** Cellular-viability assays demonstrated that silencing *SPP1* inhibited the growth of U87 and LN229 cells. **(D)** Representative images of cellular-proliferation assays using EdU staining (left) and quantification of EdU$^+$ cells (right). Nuclei were counterstained with Hoechst 33,342 (scale bar: 50 μm). **(E)** Matrigel assay demonstrated that knockdown of *SPP1* inhibited U87 and LN229 invasion (scale bar: 100 μm).

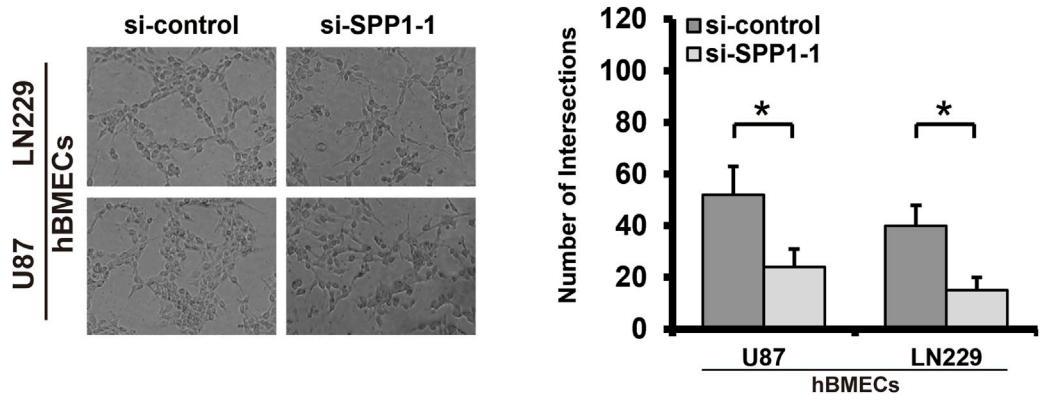


FIGURE 13
Tube formation assay. Knockdown of *SPP1* inhibited tumor angiogenesis *in vitro*. All experiments were performed in triplicate.

patients, and it might itself constitute a potential therapeutic target.

Conclusion

In summary, the study analyzed the expression pattern and predictive value of ARGs in gliomas. Furthermore, we used a risk model based on the expression of ARGs to predict survival, and the risk score was correlated with the TIME in gliomas. The risk score can be used as an independent prognostic indicator. However, further studies using prospective, large-scale, multicenter clinical cohorts are needed to validate the risk 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 authors.

Author contributions

TH, YW, and SH designed the study. XW, RW, and YS searched, collected, and pre-processed data. TH, LZ, and SH analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the Liaoning Revitalization Talents Program (No. XLYC1807253), the Shenyang Young Scientific and Technological Innovation Project (No. RC200610), and the National Natural Science Foundation of China (No. 81772653).

Acknowledgments

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript and for the free use of the TCGA and CGGA databases.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.934683/full#supplementary-material

References

Abou Khouzam, R., Brodaczewska, K., Filipiak, A., Zeinelabdin, N. A., Buart, S., Szczylik, C., et al. (2020). Tumor hypoxia regulates immune escape/invasion: Influence on angiogenesis and potential impact of hypoxic biomarkers on cancer therapies. *Front. Immunol.* 11, 613114. doi:10.3389/fimmu.2020.613114

Aldape, K., Brindle, K. M., Chesler, L., Chopra, R., Gajjar, A., Gilbert, M. R., et al. (2019). Challenges to curing primary brain tumours. *Nat. Rev. Clin. Oncol.* 16 (8), 509–520. doi:10.1038/s41571-019-0177-5

Biterge-Sut, B. (2020). A comprehensive analysis of the angiogenesis-related genes in glioblastoma multiforme vs. brain lower grade glioma. *Arq. Neuropsiquiatr.* 78 (1), 34–38. doi:10.1590/0004-282X20190131

Braun, D. A., Burke, K. P., and Van Allen, E. M. (2016). Genomic approaches to understanding response and resistance to immunotherapy. *Clin. Cancer Res.* 22 (23), 5642–5650. doi:10.1158/1078-0432.CCR-16-0066

Butler, M., Pongor, L., Su, Y. T., Xi, L., Raffeld, M., Quezado, M., et al. (2020). MGMT status as a clinical biomarker in glioblastoma. *Trends Cancer* 6 (5), 380–391. doi:10.1016/j.trecan.2020.02.010

Carmeliet, P. (2005). Angiogenesis in life, disease and medicine. *Nature* 438 (7070), 932–936. doi:10.1038/nature04478

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann. Oncol.* 30 (1), 44–56. doi:10.1093/annonc/mdy495

Chao, B., Jiang, F., Bai, H., Meng, P., Wang, L., and Wang, F. (2022). Predicting the prognosis of glioma by pyroptosis-related signature. *J. Cell. Mol. Med.* 26 (1), 133–143. doi:10.1111/jcmm.17061

Chen, J., Hou, C., Zheng, Z., Lin, H., Lv, G., and Zhou, D. (2019). Identification of secreted phosphoprotein 1 (SPP1) as a prognostic factor in lower-grade gliomas. *World Neurosurg.* 130, e775–e785. doi:10.1016/j.wneu.2019.06.219

Chen, R., Smith-Cohn, M., Cohen, A. L., and Colman, H. (2017). Glioma subclassifications and their clinical significance. *Neurotherapeutics* 14 (2), 284–297. doi:10.1007/s13311-017-0519-x

Chen, R., Zhang, H., Wu, W., Li, S., Wang, Z., Dai, Z., et al. (2022). Antigen presentation machinery signature-derived CALR mediates migration, polarization of macrophages in glioma and predicts immunotherapy response. *Front. Immunol.* 13, 833792. doi:10.3389/fimmu.2022.833792

Chen, Z., Wu, H., Yang, H., Fan, Y., Zhao, S., and Zhang, M. (2021a). Identification and validation of RNA-binding protein-related gene signature revealed potential associations with immunosuppression and drug sensitivity in glioma. *Cancer Med.* 10 (20), 7418–7439. doi:10.1002/cam4.4248

Chen, Z., Wu, T., Yan, Z., and Zhang, M. (2021b). Identification and validation of an 11-ferroptosis related gene signature and its correlation with immune checkpoint molecules in glioma. *Front. Cell Dev. Biol.* 9, 652599. doi:10.3389/fcell.2021.652599

Cloughesy, T. F., Mochizuki, A. Y., Orpilla, J. R., Hugo, W., Lee, A. H., Davidson, T. B., et al. (2019). Neoadjuvant anti-PD-1 immunotherapy promotes a survival benefit with intratumoral and systemic immune responses in recurrent glioblastoma. *Nat. Med.* 25 (3), 477–486. doi:10.1038/s41591-018-0337-7

Cong, P., Wu, T., Huang, X., Liang, H., Gao, X., Tian, L., et al. (2021). Identification of the role and clinical prognostic value of target genes of m6A RNA methylation regulators in glioma. *Front. Cell Dev. Biol.* 9, 709022. doi:10.3389/fcell.2021.709022

Coniglio, S. J., and Segall, J. E. (2013). Review: Molecular mechanism of microglia stimulated glioblastoma invasion. *Matrix Biol.* 32 (7-8), 372–380. doi:10.1016/j.matbio.2013.07.008

Crocker, M., Ashley, S., Giddings, I., Petrik, V., Hardcastle, A., Aherne, W., et al. (2011). Serum angiogenic profile of patients with glioblastoma identifies distinct tumor subtypes and shows that TIMP-1 is a prognostic factor. *Neuro. Oncol.* 13 (1), 99–108. doi:10.1093/neuonc/noq170

Deng, X., Lin, D., Zhang, X., Shen, X., Yang, Z., Yang, L., et al. (2020). Profiles of immune-related genes and immune cell infiltration in the tumor microenvironment of diffuse lower-grade gliomas. *J. Cell. Physiol.* 235 (10), 7321–7331. doi:10.1002/jcp.29633

Eckel-Passow, J. E., Lachance, D. H., Molinaro, A. M., Walsh, K. M., Decker, P. A., Sicotte, H., et al. (2015). Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N. Engl. J. Med.* 372 (26), 2499–2508. doi:10.1056/NEJMoa1407279

Fan, Y., Peng, X., Wang, Y., Li, B., and Zhao, G. (2021). Comprehensive analysis of HDAC family identifies HDAC1 as a prognostic and immune infiltration indicator and HDAC1-related signature for prognosis in glioma. *Front. Mol. Biosci.* 8, 720020. doi:10.3389/fmolb.2021.720020

Feng, E., Liang, T., Wang, X., Du, J., Tang, K., Wang, X., et al. (2019). Correlation of alteration of HLA-F expression and clinical characterization in 593 brain glioma samples. *J. Neuroinflammation* 16 (1), 33. doi:10.1186/s12974-019-1418-3

Furnari, F. B., Fenton, T., Bachoo, R. M., Mukasa, A., Stommel, J. M., Stegh, A., et al. (2007). Malignant astrocytic glioma: Genetics, biology, and paths to treatment. *Genes Dev.* 21 (21), 2683–2710. doi:10.1101/gad.1596707

Gao, F., Wang, Z., Gu, J., Zhang, X., and Wang, H. (2021). A hypoxia-associated prognostic gene signature risk model and prognosis predictors in gliomas. *Front. Oncol.* 11, 726794. doi:10.3389/fonc.2021.726794

Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W., Kim, D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* 21 (8), 938–945. doi:10.1038/nm.3909

Guo, X., Pan, Y., Xiong, M., Sanapala, S., Anastasaki, C., Cobb, O., et al. (2020). Midkine activation of CD8(+) T cells establishes a neuron-immune-cancer axis responsible for low-grade glioma growth. *Nat. Commun.* 11 (1), 2177. doi:10.1038/s41467-020-15770-3

Han, S., Li, X., Qiu, B., Jiang, T., and Wu, A. (2015). Can lateral ventricle contact predict the ontogeny and prognosis of glioblastoma? *J. Neurooncol.* 124 (1), 45–55. doi:10.1007/s11060-015-1818-x

Han, S., Wang, C., Qin, X., Xia, J., and Wu, A. (2017). LPS alters the immunophenotype of glioma and glioma stem-like cells and induces *in vivo* antitumor immunity via TLR4. *J. Exp. Clin. Cancer Res.* 36 (1), 83. doi:10.1186/s13046-017-0552-y

He, J., Jiang, Z., Lei, J., Zhou, W., Cui, Y., Luo, B., et al. (2022). Prognostic value and therapeutic perspectives of CXCR members in the glioma microenvironment. *Front. Genet.* 13, 787141. doi:10.3389/fgene.2022.787141

Hicklin, D. J., and Ellis, L. M. (2005). Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. *J. Clin. Oncol.* 23 (5), 1011–1027. doi:10.1200/JCO.2005.06.081

Hu, Y., Song, J., Wang, Z., Kan, J., Ge, Y., Wang, D., et al. (2021). A novel S100 family-based signature associated with prognosis and immune microenvironment in glioma. *J. Oncol.* 2021, 3586589. doi:10.1155/2021/3586589

Jin, X., Nie, E., Zhou, X., Zeng, A., Yu, T., Zhi, T., et al. (2017). Fstl1 promotes glioma growth through the BMP4/smad1/5/8 signaling pathway. *Cell. Physiol. biochem.* 44 (4), 1616–1628. doi:10.1159/000485759

Kennedy, B. C., Showers, C. R., Anderson, D. E., Anderson, L., Canoll, P., Bruce, J. N., et al. (2013). Tumor-associated macrophages in glioma: Friend or foe? *J. Oncol.* 2013, 486912. doi:10.1155/2013/486912

Kim, H. S., Chang, C. Y., Yoon, H. J., Kim, K. S., Koh, H. S., Kim, S. S., et al. (2020). Glial TIM-3 modulates immune responses in the brain tumor microenvironment. *Cancer Res.* 80 (9), 1833–1845. doi:10.1158/0008-5472.CAN-19-2834

Kim, H., Zheng, S., Amini, S. S., Virk, S. M., Mikkelsen, T., Brat, D. J., et al. (2015). Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* 25 (3), 316–327. doi:10.1101/gr.180612.114

Kruger, S., Ilmer, M., Kobold, S., Cadilha, B. L., Endres, S., Ormanns, S., et al. (2019). Advances in cancer immunotherapy 2019 – latest trends. *J. Exp. Clin. Cancer Res.* 38 (1), 268. doi:10.1186/s13046-019-1266-0

Li, C., Liu, F., Sun, L., Liu, Z., and Zeng, Y. (2022a). Natural killer cell-related gene signature predicts malignancy of glioma and the survival of patients. *BMC Cancer* 22 (1), 230. doi:10.1186/s12885-022-09230-y

Li, H., Wang, G., Wang, W., Pan, J., Zhou, H., Han, X., et al. (2021). A focal adhesion-related gene signature predicts prognosis in glioma and correlates with radiation response and immune microenvironment. *Front. Oncol.* 11, 698278. doi:10.3389/fonc.2021.698278

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). Timer: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.CAN-17-0307

Li, W., and Graeber, M. B. (2012). The molecular profile of microglia under the influence of glioma. *Neuro. Oncol.* 14 (8), 958–978. doi:10.1093/neuonc/nos116

Li, X., Wang, Y., Wu, W., Xiang, J., Qi, L., Wang, N., et al. (2022b). A novel risk score model based on eleven extracellular matrix-related genes for predicting overall survival of glioma patients. *J. Oncol.* 2022, 4966820. doi:10.1155/2022/4966820

Li, X., Wang, Y., Wu, W., Xiang, J., Wang, M., and Yu, H. (2022c). A novel DNA damage and repair-related gene signature to improve predictive capacity of overall survival for patients with gliomas. *J. Cell. Mol. Med.* 26, 3736–3750. doi:10.1111/jcmm.17406

Lindau, D., Gielen, P., Kroesen, M., Wesseling, P., and Adema, G. J. (2013). The immunosuppressive tumour network: Myeloid-derived suppressor cells, regulatory

T cells and natural killer T cells. *Immunology* 138 (2), 105–115. doi:10.1111/imm.12036

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 world Health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* 131 (6), 803–820. doi:10.1007/s00401-016-1545-1

Luoto, S., Hermelo, I., Vuorinen, E. M., Hannus, P., Kesseli, J., Nykter, M., et al. (2018). Computational characterization of suppressive immune microenvironments in glioblastoma. *Cancer Res.* 78 (19), 5574–5585. doi:10.1158/0008-5472.CAN-17-3714

Machulla, H., Steinborn, F., Schaaf, A., Heidecke, V., and Rainov, N. (2001). Brain glioma and human leukocyte antigens (HLA)--is there an association. *J. Neurooncol.* 52 (3), 253–261. doi:10.1023/a:1010612327647

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37 (7), 773–782. doi:10.1038/s41587-019-0114-2

Onishi, M., Ichikawa, T., Kurozumi, K., and Date, I. (2011). Angiogenesis and invasion in glioma. *Brain Tumor Pathol.* 28 (1), 13–24. doi:10.1007/s10014-010-0007-z

Ostrom, Q. T., Cioffi, G., Gittleman, H., Patil, N., Waite, K., Kruchko, C., et al. (2019). CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2012-2016. *Neuro. Oncol.* 21 (5), v1–v100. doi:10.1093/neuonc/noz150

Peng, C., Chen, H., Li, Y., Yang, H., Qin, P., Ma, B., et al. (2021). LRIG3 suppresses angiogenesis by regulating the PI3K/AKT/VEGFA signaling pathway in glioma. *Front. Oncol.* 11, 621154. doi:10.3389/fonc.2021.621154

Pereira, M. B., Barros, L. R. C., Bracco, P. A., Vigo, A., Boroni, M., Bonamino, M. H., et al. (2018). Transcriptional characterization of immunological infiltrates and their relation with glioblastoma patients overall survival. *Oncoimmunology* 7 (6), e1431083. doi:10.1080/2162402X.2018.1431083

Qi, Y., Yang, X., Ji, C., Tang, C., and Xie, L. (2022). Identification of an IL-4-related gene risk signature for malignancy, prognosis and immune phenotype prediction in glioma. *Brain Sci.* 12 (2), 181. doi:10.3390/brainsci12020181

Quail, D. F., and Joyce, J. A. (2017). The microenvironmental landscape of brain tumors. *Cancer Cell* 31 (3), 326–341. doi:10.1016/j.ccell.2017.02.009

Reddy, S. P., Britto, R., Vinnakota, K., Aparna, H., Sreepathi, H. K., Thota, B., et al. (2008). Novel glioblastoma markers with diagnostic and prognostic value identified through transcriptome analysis. *Clin. Cancer Res.* 14 (10), 2978–2987. doi:10.1158/1078-0432.CCR-07-4821

Ren, H., Zhu, J., Yu, H., Bazhin, A. V., Westphalen, C. B., Renz, B. W., et al. (2020). Angiogenesis-related gene expression signatures predicting prognosis in gastric cancer patients. *Cancers (Basel)* 12 (12), E3685. doi:10.3390/cancers12123685

Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Tamura, R., Tanaka, T., Akasaki, Y., Murayama, Y., Yoshida, K., and Sasaki, H. (2019). The role of vascular endothelial growth factor in the hypoxic and immunosuppressive tumor microenvironment: Perspectives for therapeutic implications. *Med. Oncol.* 37 (1), 2. doi:10.1007/s12032-019-1329-2

Tan, Z., Chen, K., Wu, W., Zhou, Y., Zhu, J., Wu, G., et al. (2018). Overexpression of HOXC10 promotes angiogenesis in human glioma via interaction with PRMT5 and upregulation of VEGFA expression. *Theranostics* 8 (18), 5143–5158. doi:10.7150/thno.27310

Tian, Y., Ke, Y., and Ma, Y. (2020). High expression of stromal signatures correlated with macrophage infiltration, angiogenesis and poor prognosis in glioma microenvironment. *PeerJ* 8, e9038. doi:10.7717/peerj.9038

Uneda, A., Kurozumi, K., Fujimura, A., Fujii, K., Ishida, J., Shimazu, Y., et al. (2021). Differentiated glioblastoma cells accelerate tumor progression by shaping the tumor microenvironment via CCN1-mediated macrophage infiltration. *Acta Neuropathol. Commun.* 9 (1), 29. doi:10.1186/s40478-021-01124-7

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131 (4), 281–285. doi:10.1007/s12064-012-0162-3

Wang, G., Hu, J.-Q., Liu, J.-Y., and Zhang, X.-M. (2022). Angiogenesis-related gene signature-derived risk score for glioblastoma: Prospects for predicting prognosis and immune heterogeneity in glioblastoma. *Front. Cell Dev. Biol.* 10, 778286. doi:10.3389/fcell.2022.778286

Wang, X., Zhang, L., Song, Y., Jiang, Y., Zhang, D., Wang, R., et al. (2021). MCM8 is regulated by EGFR signaling and promotes the growth of glioma stem cells through its interaction with DNA-replication-initiating factors. *Oncogene* 40 (27), 4615–4624. doi:10.1038/s41388-021-01888-1

Weenink, B., Draaisma, K., Ooi, H. Z., Kros, J. M., Sillevis Smitt, P. A. E., Debets, R., et al. (2019). Low-grade glioma harbors few CD8 T cells, which is accompanied by decreased expression of chemo-attractants, not immunogenic antigens. *Sci. Rep.* 9 (1), 14643. doi:10.1038/s41598-019-51063-6

Wick, W., Weller, M., van den Bent, M., Sanson, M., Weiler, M., von Deimling, A., et al. (2014). MGMT testing--the challenges for biomarker-based glioma treatment. *Nat. Rev. Neurol.* 10 (7), 372–385. doi:10.1038/nrneurol.2014.100

Xiao, S., Yu, J., Yuan, X., and Chen, Q. (2022). Identification of a tripartite motif family gene signature for predicting the prognosis of patients with glioma. *Am. J. Transl. Res.* 14 (3), 1535–1550.

Xu, S., Tang, L., Dai, G., Luo, C., and Liu, Z. (2020a). Expression of m6A regulators correlated with immune microenvironment predicts therapeutic efficacy and prognosis in gliomas. *Front. Cell Dev. Biol.* 8, 594112. doi:10.3389/fcell.2020.594112

Xu, S., Tang, L., Dai, G., Luo, C., and Liu, Z. (2021). Immune-related genes with APA in microenvironment indicate risk stratification and clinical prognosis in grade II/III gliomas. *Mol. Ther. Nucleic Acids* 23, 1229–1242. doi:10.1016/j.omtn.2021.01.033

Xu, S., Tang, L., Li, X., Fan, F., and Liu, Z. (2020b). Immunotherapy for glioma: Current management and future application. *Cancer Lett.* 476, 1–12. doi:10.1016/j.canlet.2020.02.002

Xu, Y., Li, R., Li, X., Dong, N., Wu, D., Hou, L., et al. (2020c). An autophagy-related gene signature associated with clinical prognosis and immune microenvironment in gliomas. *Front. Oncol.* 10, 571189. doi:10.3389/fonc.2020.571189

Yan, Z., Gao, Y., Yu, J., Shen, Z., and Bu, X. (2022). Identification of an inflammatory response-related gene signature to predict survival and immune status in glioma patients. *J. Immunol. Res.* 2022, 8972730. doi:10.1155/2022/8972730

Yang, Z., Chen, Z., Wang, Y., Wang, Z., Zhang, D., Yue, X., et al. (2022). A novel defined pyroptosis-related gene signature for predicting prognosis and treatment of glioma. *Front. Oncol.* 12, 717926. doi:10.3389/fonc.2022.717926

Yin, W., Jiang, X., Tan, J., Xin, Z., Zhou, Q., Zhan, C., et al. (2020). Development and validation of a tumor mutation burden-related immune prognostic model for lower-grade glioma. *Front. Oncol.* 10, 1409. doi:10.3389/fonc.2020.01409

Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Zhai, L., Ladomersky, E., Lauing, K. L., Wu, M., Genet, M., Gritsina, G., et al. (2017). Infiltrating T cells increase Ido1 expression in glioblastoma and contribute to decreased patient survival. *Clin. Cancer Res.* 23 (21), 6650–6660. doi:10.1158/1078-0432.CCR-17-0120

Zhang, C., Guo, L., Su, Z., Luo, N., Tan, Y., Xu, P., et al. (2021a). Tumor immune microenvironment landscape in glioma identifies a prognostic and immunotherapeutic signature. *Front. Cell Dev. Biol.* 9, 717601. doi:10.3389/fcell.2021.717601

Zhang, M., Cheng, Y., Xue, Z., Sun, Q., and Zhang, J. (2021b). A novel pyroptosis-related gene signature predicts the prognosis of glioma through immune infiltration. *BMC Cancer* 21 (1), 1311. doi:10.1186/s12885-021-09046-2

Zhang, Y., Xie, Y., He, L., Tang, J., He, Q., Cao, Q., et al. (2021c). 1p/19q co-deletion status is associated with distinct tumor-associated macrophage infiltration in IDH mutated lower-grade gliomas. *Cell. Oncol.* 44 (1), 193–204. doi:10.1007/s13402-020-00561-1

Zhang, Y., Zhang, C., Yang, Y., Wang, G., Wang, Z., Liu, J., et al. (2022). Pyroptosis-related gene signature predicts prognosis and indicates immune microenvironment infiltration in glioma. *Front. Cell Dev. Biol.* 10, 862493. doi:10.3389/fcell.2022.862493

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 12 August 2022
DOI 10.3389/fmolb.2022.928530

Check for updates

OPEN ACCESS

EDITED BY
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

REVIEWED BY
Shuting Yang,
Second Xiangya Hospital, Central South University, China
Zeliha Selamoglu,
Niğde Ömer Halisdemir University, Turkey

*CORRESPONDENCE
Hammad Naveed,
hammad.naveed@nu.edu.pk

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 25 April 2022
ACCEPTED 04 July 2022
PUBLISHED 12 August 2022

CITATION
Saeed N and Naveed H (2022), Medical terminology-based computing system: a lightweight post-processing solution for out-of-vocabulary multi-word terms.
Front. Mol. Biosci. 9:928530.
doi: 10.3389/fmolb.2022.928530

COPYRIGHT
© 2022 Saeed and Naveed. 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.

Medical terminology-based computing system: a lightweight post-processing solution for out-of-vocabulary multi-word terms

Nadia Saeed and Hammad Naveed*

Computational Biology Research Lab, Department of Computer Science, National University of Computer and Emerging Sciences (NUCES-FAST), Islamabad, Pakistan

The linguistic rules of medical terminology assist in gaining acquaintance with rare/complex clinical and biomedical terms. The medical language follows a Greek and Latin-inspired nomenclature. This nomenclature aids the stakeholders in simplifying the medical terms and gaining semantic familiarity. However, natural language processing models misrepresent rare and complex biomedical words. In this study, we present MedTCS—a lightweight, post-processing module—to simplify hybridized or compound terms into regular words using medical nomenclature. MedTCS enabled the word-based embedding models to achieve 100% coverage and enabled the BiowordVec model to achieve high correlation scores (0.641 and 0.603 in UMNSRS similarity and relatedness datasets, respectively) that significantly surpass the n-gram and sub-word approaches of FastText and BERT. In the downstream task of named entity recognition (NER), MedTCS enabled the latest clinical embedding model of FastText-OA-All-300d to improve the F1-score from 0.45 to 0.80 on the BC5CDR corpus and from 0.59 to 0.81 on the NCBI-Disease corpus, respectively. Similarly, in the drug indication classification task, our model was able to increase the coverage by 9% and the F1-score by 1%. Our results indicate that incorporating a medical terminology-based module provides distinctive contextual clues to enhance vocabulary as a post-processing step on pre-trained embeddings. We demonstrate that the proposed module enables the word embedding models to generate vectors of out-of-vocabulary words effectively. We expect that our study can be a stepping stone for the use of biomedical knowledge-driven resources in NLP.

KEYWORDS

medical terminology, named entity recognition, linguistic approach, natural language processing, biomedical nomenclature, out-of-vocabulary

1 Introduction

Familiarity with medical terminology assists medical practitioners and other stakeholders like doctors, nurses, and clinicians to understand rare and complex vocabulary. The evolution of medical terminology presents challenges in promoting the use of electronic health and medical records. For example, most medical terms originate from Greek and Latin words, making reading and spelling difficult Henderson and Dorsey (2019); Banay (1948). Medical researchers acquire conceptual skills with thorough learning of medical terms, dictionaries, and references such as Merriam-Webster Merriam-Webster (2018), WebMD WebMD (2012), and MedicineNet MedicineNet (2007), etc.

The electronic health records (EHRs) contain the diagnoses, pharmacological, and drug-disease concepts that provide a complete view of a patient's health. EHRs can inform drug discovery, treatment pathways, and real-world safety assessments. Unstructured text from EHRs can be encoded in a structured format (vectors) for downstream analysis using NLP methods. Unfortunately, the word embedding models faced the Out-of-vocabulary (OOV) words problem or used ineffective sub-word representations that caused low performance in intrinsic tasks to retrieve conceptual properties.

Popular embedding models including BERT (Devlin et al., 2019), ELMO (Peters et al., 2018), and FastText (Bojanowski et al., 2017) solve the OOV problem by using pre-processing tokenization techniques based on WordPiece (Wu et al., 2016), characters, and n-grams. These traditional NLP approaches are not built to understand the unique vocabulary and grammar of medical texts. For example, *mastodynia* is a disease whose meaning can be approximated from related and simple words like *breast, pain, and discomfort* rather than to approximate it with its non-logical sub-words or n-grams like [CLS],mast,##ody, and ##nia [SEP].

Biomedical and clinical terms have unique and complex characteristics such as prefixes, roots, suffixes, etc., therefore requiring a more focused effort around methodologies within the medical NLP domain (Banay, 1948; Meystre et al., 2008; Cohen and Demner-Fushman, 2014; Leaman et al., 2015; Henderson and Dorsey, 2019). In recent years, biomedical and clinical embedding models such as BioWordVec (Zhang et al., 2019) and BioNLP (Chiu et al., 2016) models have been trained under low capacity resource requirements like the Gensim library (Řehřek and Sojka, 2011). However, these models generally follow the Word2Vec (Mikolov et al., 2013a; Mikolov et al., 2013b) and GloVe (Pennington et al., 2014) algorithms, which face the OOV problem. The embedding models trained using the FastText algorithm (Bojanowski et al., 2017) claim to have solved the OOV problem, however they are ineffective.

The pre-trained embedding models generate either context-sensitive or distributed representations of word vectors. The context-sensitive models generate multiple embeddings for a word that capture the context based on its positional encoding learned using transformers or recurrent neural networks (RNN). Bidirectional Encoder Representations from Transformers (BERT) is a popular embedding model (Devlin et al., 2019), that has been extended to clinical and biomedical domains [ClinicalBERT Huang et al. (2019) and BioBERT Lee et al. (2020)]. These models tackle the OOV problem with the WordPiece algorithm (Wu et al., 2016) that represents a word by its frequent sub-words, e.g., immunoglobulin → (i,mm,uno,g,lo,bul,in). Embeddings from Language Models (ELMO) is another context-sensitive model that generates word-level embeddings using multiple convolutional neural networks (CNNs) with bi-directional LSTM (BiLSTM) (Peters et al., 2018). ELMO has also been extended to generate biomedical and clinical embeddings (Zhu et al., 2018; Jin et al., 2019; Subramanyam and Sangeetha. 2020). These studies deal with the OOV problem through character-level embeddings. Boukkouri et al. showed that character-level embedding was a better approach to removing biases in sub-words for biomedical terms than WordPiece e.g., choledocholithiasis → (cho,led,och,oli,thi,asi,s) (Boukkouri et al., 2020; Wu et al., 2016). The context-sensitive models are expensive, both in terms of computational and space resources since they train millions of hyperparameters with multiple attention heads.

The distributed representation models learn embeddings based on the word usage in a given corpus. The resultant vectors capture the contextual similarity between words. These static models generate a single vector per word and are trained either under Word2Vec (Mikolov et al., 2013b), GloVe (Pennington et al., 2014), or FastText (Bojanowski et al. (2017)). FastText enriches each word vector with its respective n-grams. It handles the OOV problem by leveraging the sum of n-gram vectors of the unknown word, e.g., n = 3, myocarditis → < my, myo, yoc, oca, car, ard, rdi, dit, iti, tis, is > . On the other hand, the embedding models trained by Word2Vec and GloVe face the OOV problem. These models replace unknown words with tags such as < UNK > or a randomly generated vector, where different unknown words lose their uniqueness.

In this study, we proposed MedTCS, a novel medical terminology-based module that assists the pre-trained embedding models to generate vectors for unknown words and compound terms. It is an innovative post-processing solution that explores the given search space for those terms that are not directly present but whose semantic information is. MedTCS turns the word into its meaningful sub-words using the biomedical segmentation model. Ultimately, MedTCS helps the distributed representation models handle the OOV problem effectively.

We have compared MedTCS with recent state-of-the-art embedding models to investigate the effectiveness of capturing semantic information without encountering OOV problems. Our results showed that MedTCS enhanced the performance of pre-

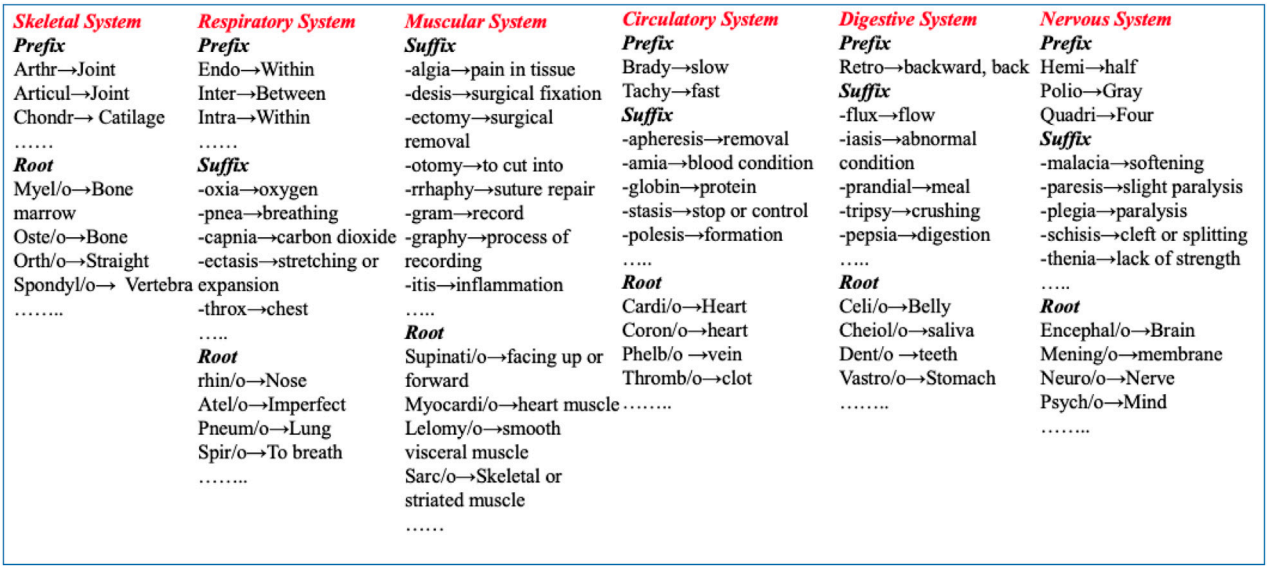


FIGURE 1
Understanding biomedical terms by mapping term components to human organ system.

trained models significantly in terms of coverage and/or semantic correlation. Moreover, we conducted experiments to assess the usefulness of enriched embedded vectors for downstream NER tasks (disease name identification and drug indication classification). MedTCS performed better than FastText in terms of performance and resource consumption on all tasks. The MedTCS module enhanced the performance of the FastText word vectors as compared to the n-gram and sub-word approaches used for unknown words (Flamholz et al., 2022). Furthermore, MedTCS is extensible with new terminologies and content.

2 Methodology

MedTCS is a lightweight module implemented in Python. It is a knowledge-driven system for forming terms by pluralizing, singularizing, and deconstructing words.

2.1 Meta-data collection

2.1.1 Word component dictionary

In MedTCS, we build meta-dictionaries for the prefixes, roots, and suffixes defining the meanings of medical term components. In addition to the lexical normalization and plural conversion of the unknown term, we have developed medical terminology-based look-up dictionaries for the parser by collecting information from "Medical Terminology for Dummies" (Henderson and Dorsey, 2019). The three semantic dictionaries contain 467 root words, 432 prefixes, and 112 suffixes, along with their corresponding meanings as shown in Figure 1.

2.1.2 Word segmenter model

MedTCS used Morfessor as a word segmenter model (Smit et al., 2014). In order to train the semi-supervised Morfessor 2.0 model, we used a corpora of 240 k words consisting of medical academic word list, e-biology, e-chemistry, and NLTK words (Bird and Loper, 2004; Wang et al., 2008).

2.2 MedTCS framework

Figure 2 provides a high-level description of our MedTCS module to encode OOV words from a set of sentences or words. In step (a), the OOV words are normalised for multiple morphological rules (represented as $N^{r}_{1,\ldots,n}$). In step (b), the remaining OOV words are exchanged with its plural or singular form by applying medical terminology-based rules (represented as $R^{r}_{1,\ldots,n}$). At each step, the normalized terms are encoded into vectors. In the succeeding steps (c) and (d), the words are passed to the parser, where dictionaries of prefix p, root r, and suffix s are used to tokenize them (represented as $P^{p,r,s}_{1,\ldots,n}$). Each component of the term is replaced with their respective meaning in the dictionary as a word list (represented as $M^{p_{w_i..w_k},r_{w_i..w_k},s_{w_i..w_k}}_{1,\ldots,n}$). The encoder encodes the tokens into its mean vector. Finally, the remaining non-encoded words are

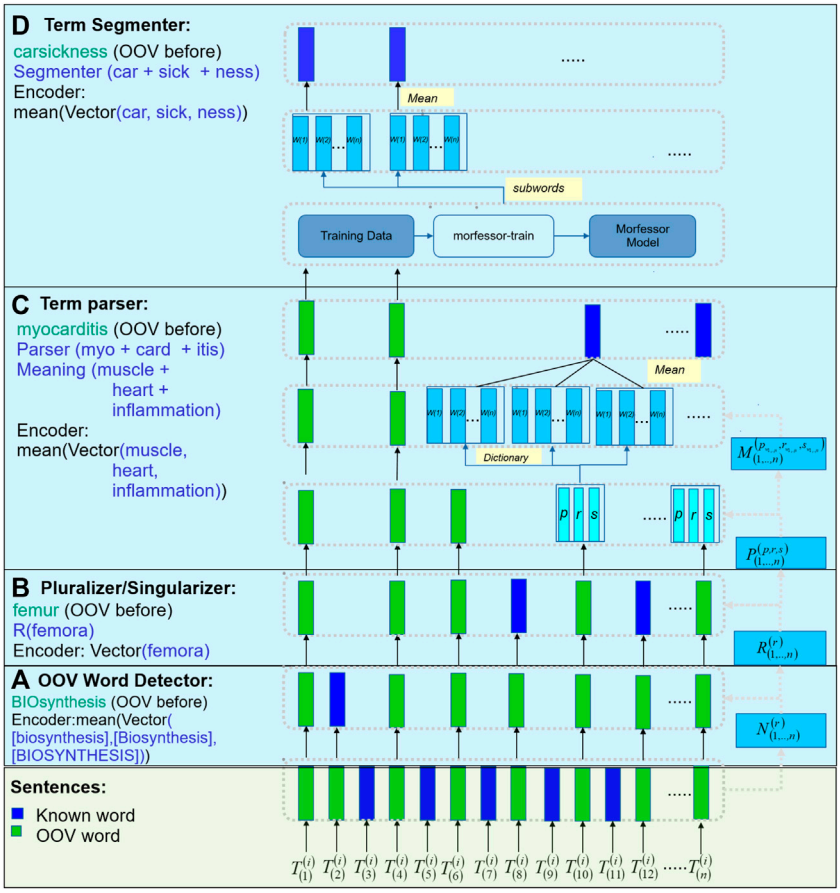


FIGURE 2
MedTCS framework: **(A)** MedTCS detector normalizes the unknown terms and search in vocabulary; **(B)** Rule-based pluralizer or singularizer sub-module used to normalize the unknown terms; **(C)** Architecture for term-parser, where the compound words encode for its components that infer from the dictionary for its semantic words that encode as its mean vector; **(D)** Architecture for term segmenter, a pre-trained segmentation model segments the word into subwords that encodes as its mean vector.

TABLE 1 Statistics of Datasets.

Evaluation	Dataset	Corpus size	Type
Intrinsic Evaluation	UMNSRS-similarity Pakhomov et al. (2010)	566 term pairs	Pairwise similarity
	UMNSRS-relatedness Pakhomov et al. (2010)	588 term pairs	Pairwise relatedness
	MyoSRS Pakhomov et al. (2011)	101 term pairs	Pairwise relatedness
	EHR-RelB Schulz et al. (2020)	3630 term pairs	Pairwise relatedness
Extrinsic Evaluation	Dataset		
	BC5CDR Wang et al. (2019)	1500 articles	Disease Name
	NCBI-Disease Wang et al. (2019)	793 abstracts	Disease Name
	DICE Bhatt et al. (2021)	7231 sentences	Drug Indication

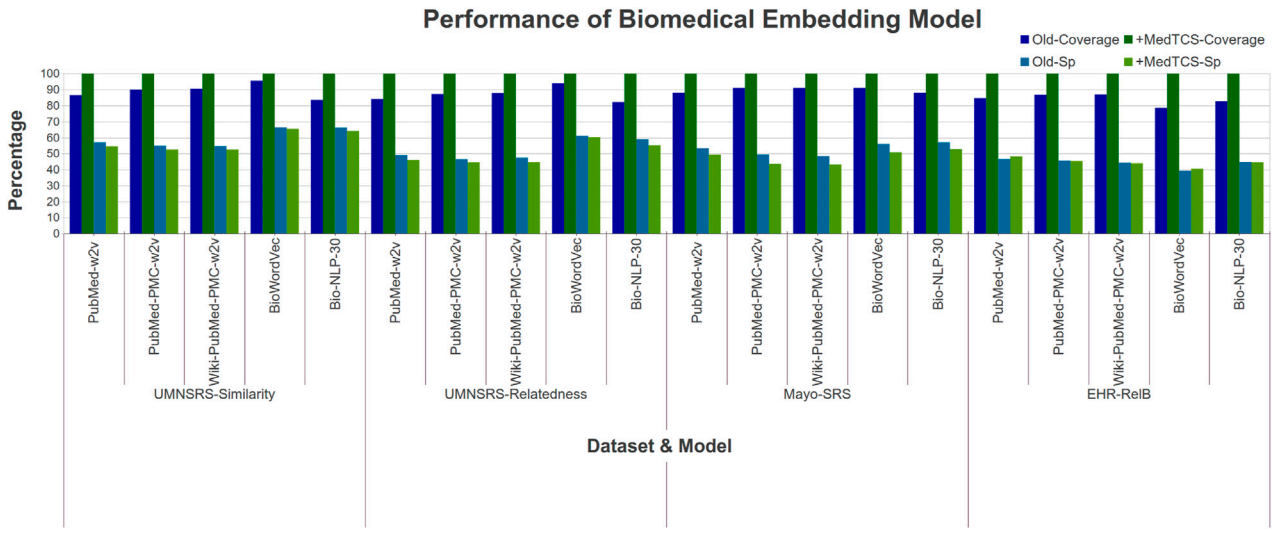


FIGURE 3
Comparison of performance variations in biomedical embedding model after adding MedTCS module on datasets of Table 1 for intrinsic evaluation.

passed to the pre-trained term segmenter model to intra-tokenize into meaningful words (that are also encoded as mean vectors).

2.2.1 MedTCS OOV word detector

The MedTCS OOV word detector identifies whether a token is known or unknown for a given vocabulary. The unrecognized word is passed through multiple normalization steps: 1) lexical property of the alphabetic case is applied, 2) intra-term punctuation marks are retained while ignoring starting and ending symbols, and 3) apostrophe symbols for OOV word detection are normalized.

2.2.2 MedTCS pluralizer/singularizer

The MedTCS pluralizer is based on the plural rules defined in medical terminology and implemented as a finite state machine. The sigularizer acts as a reverse finite state machine of the pluralizer.

2.2.3 MedTCS term parser

The MedTCS term parser was applied to an unknown word in two parts. First, the rule-based parser breaks the word into components of medical terminology, i.e., root, prefix, and suffix. Second, this parser implemented a dictionary lookup algorithm on each component to map its meaning. These dictionaries contained the definitions of the components of the medical terms collected from medical notes (Banay, 1948; Cohen and Demner-Fushman, 2014; Henderson and Dorsey, 2019). Each component in the dictionary belonged to one of the following human organ systems as shown in Figure 1 e.g., -pnea → breathing was a suffix belonging to the respiratory system. The root component is normalized for its combined form, like pneum/o → lung. Each component incrementally contributes in generating the vector representation of the unknown word. Each discovered vector by MedTCS term parser belonged to the lexical part of the unknown word and had attributes defined in the medical terminology. For example, *choledocholithiasis* → ["choledoch" (prefix)]+["o"]+["lithiasis" (suffix)] → [common bile duct]+[calculus or stone]. In case the term parser does not return a valid vector, the term segmenter was executed to determine meaningful sub-words of the unknown word.

2.2.4 MedTCS term segmenter

The MedTCS term segmenter is a wrapper around the Morfessor 2.0 module to acquire the meaningful sub-word units of an unknown term (Virpioja et al., 2013; Smit et al., 2014). We trained the system on a subset of Biology, Chemistry, and English corpora. Our word-level segmentation system returned the average vector of meaningful sub-words of an unknown term (like *seasickness* → sea + sick + ness).

2.3 Datasets

In addition to the widely tested UMNSRS similarity and relatedness datasets (Pakhomov et al., 2010), and the MyoSRS dataset (Pakhomov et al., 2011), our intrinsic evaluation included the latest and comparatively large benchmark named the EHR-RelB dataset Schulz et al. (2020). These datasets consist of word

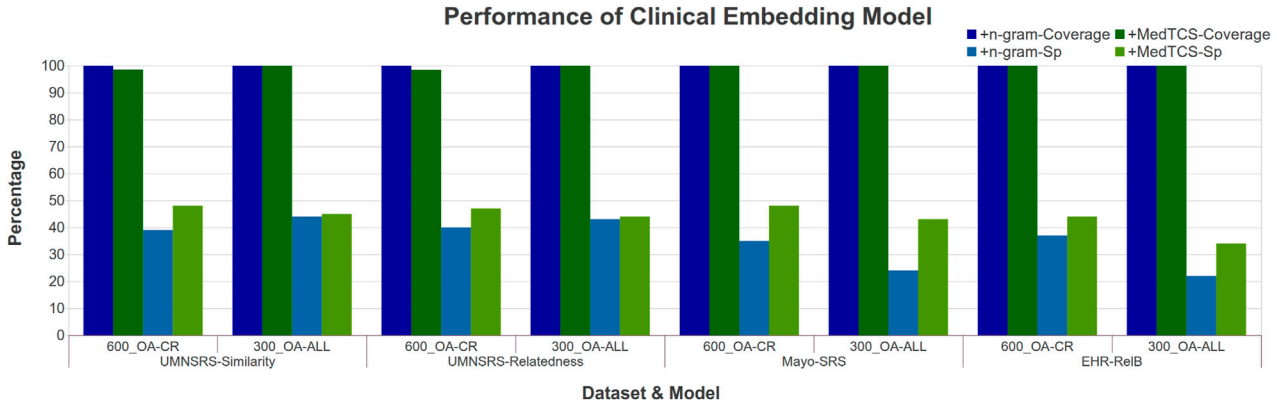


FIGURE 4
Comparison of performance variations in clinical embedding model after adding MedTCS module on datasets of Table 1 for intrinsic evaluation.

TABLE 2 Comparison of sub-word embeddings with word embedding + MedTCS on the UMNSRS-Similarity datasets.

Model		Version	Sp
BERT	BERT Devlin et al. (2019)	bert-base-uncased	0.07
	BioBert Lee et al. (2020)	dmis-lab/biobert-v1.1	0.30
	BlueBert Peng et al. (2019)	bionlp/bluebert_pubmed_mimic_uncased_L-12_H-768_A-12	0.36
	Bio_ClinicalBERT	emilyalsentzer/Bio_ClinicalBERT	0.23
Model	Alsentzer et al. (2019)	allenai/scibert_scivocab_uncased	0.18
	SciBERT Beltagy et al. (2019)		
	PubMedBERT Gu et al.(2022)	microsoft/BiomedNLP-PubMed BERT-base-uncased-abstract-fulltext	0.23
	CODER Yuan et al. (2022)	GanjinZero/UMLSBert_ENG	0.47
Word2Vec	PubMed-w2v	PubMed-w2v.bin	0.52
	+MedTCS		
	PubMed-PMC-w2v	//	0.49
Model	+MedTCS		
	Wiki-PubMed-PMC-w2v	//	0.49
+	+MedTCS		
	Bio-NLP-30 Chiu et al. (2016)	Bio-NLP-30	0.63
MedTCS	+MedTCS		
	BioWordVec Zhang et al. (2019)	BioWordVec	0.64
	+MedTCS		

pairs with their similarity or relatedness scores assigned by medical experts.

We checked the applicability of the MedTCS module to extract disease names from two publicly available datasets [NCBI-Disease and BC5CDR-Disease, Wang et al. (2019)] using the *BIO* scheme. *BIO* is used to encode entity annotations as token tags, where *B* indicates the beginning of the phrase, *I* is the element within the phrase, and *O* is the element outside of the phrase. Table 1 gives the details of benchmark datasets used for performance evaluation. We also used the Drug Indication Classification and Encyclopedia (DICE) dataset Bhatt et al. (2021) to check the performance enhancement achieved by MedTCS on classifying a sentence into indication or non-indication defined for five categories (indications, contradictions, side effects, usage instructions, and clinical observations). The dataset contained 7,231 sentences that were categorized into 4,297 indications, 1,673 clinical observations,

TABLE 3 Comparison of the word embedding + MedTCS best scores with latest reported results.

Model	UMNSRS-Similarity		UMNSRS-Relatedness		Model description
	#566	Sp	#587	Sp	
BioWordVec+ Graph Embeddings (GCN) Mao and Fung (2020)	480	0.629	473	0.590	A combined model of Graph convolutional network (GCN) a path-based graph embedding with BioWordVec embedding
Context2Vec+ BioWordVec+ PubMed + PMC Singh and Jin (2020)	471	0.634	484	0.561	Composite model of contextual embedding with BioWordVec concatenated with PubMed and PMC word embedding to achieve these results
CoderBERT Kalyan and Sangeetha (2021)	543	0.543	564	0.473	A BERT-based model obtained by fine-tuned a pre-trained BioBERT on UMLS synonyms and relations
SapBERT-S Kalyan and Sangeetha (2021)	543	0.585	564	0.505	A BERT-based model fine-tuned a pre-trained PubMedBERT on UMLS using a self-alignment objective to cluster the term concept
BioWordVec +MedTCS	**566**	**0.641**	**587**	**0.603**	BioWordVec with our composed MedTCS module, to extract the vector representation of a known and unknown term

Results with highest values of correlation and coverage scores are shown in bold.

701 contraindications, 492 usage instructions, and 68 side effects.

All the datasets discussed in Table 1 are publicly available in split form.

2.4 Evaluation metrics

In NLP, intrinsic evaluation extracts the semantic properties of pre-determined ground truth concepts with encoded vectors. On the other hand, extrinsic evaluation decodes the encoded information of embedding models and evaluates their efficiency in performing downstream tasks like NER. For the extrinsic evaluation, the coverage percentage is based on the number of encoded tokens of a dataset with the respective embedding model.

$$cosine_similarity(A,B) = \frac{A.B}{\|A\|\times\|B\|}, \quad (1)$$

$$recall = \frac{TruePositives}{TruePositives + FalseNegatives}, \quad (2)$$

$$precision = \frac{TruePositives}{TruePositives + FalsePositives}, \quad (3)$$

$$F1_score = \frac{2 \times (precision \times recall)}{precision + recall}. \quad (4)$$

In intrinsic evaluation, the similarity scores are computed between the encoded term pairs using the cosine similarity as given in Eq. 1. Furthermore, these similarity scores are used with

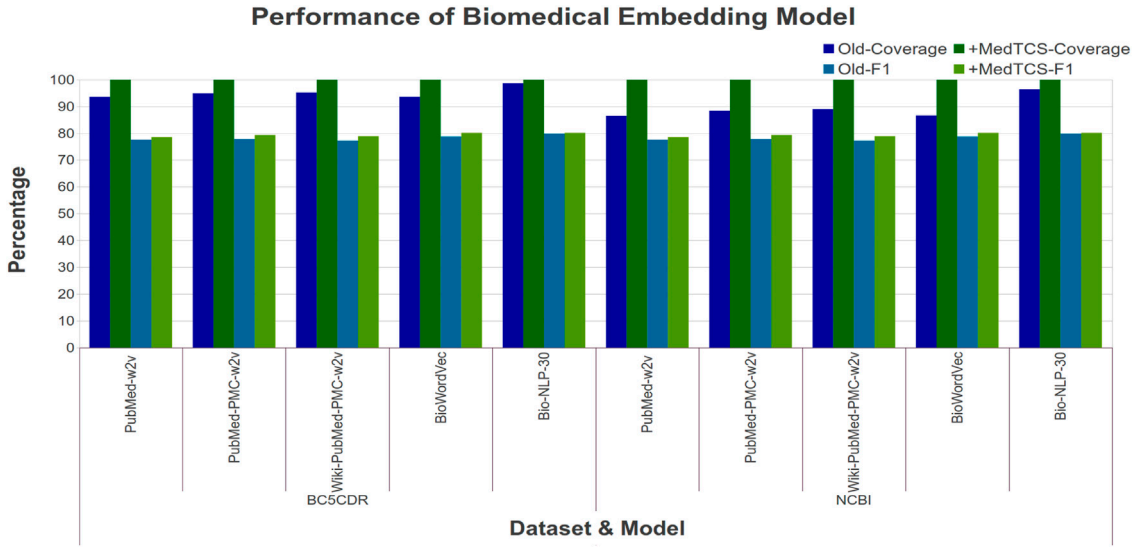


FIGURE 5
Comparison of performance variations in biomedical word embedding model after adding MedTCS module on datasets of Table 1 for NER task.

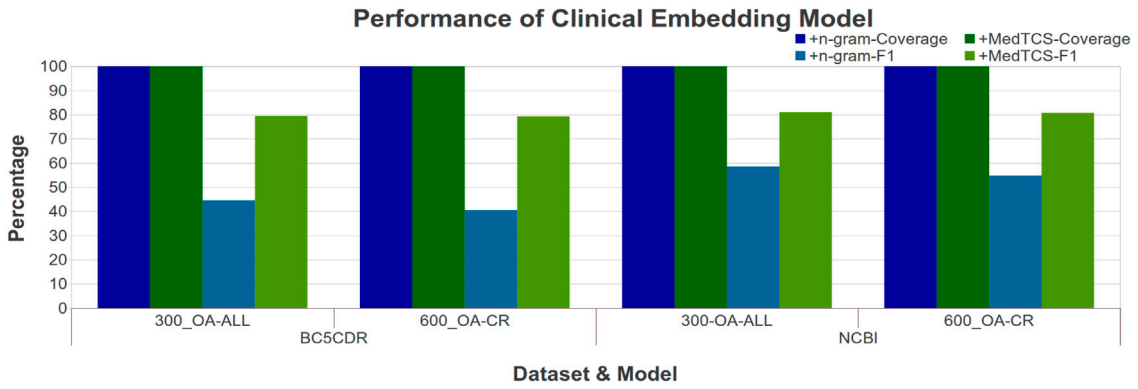


FIGURE 6
Comparison of performance variations in clinical FastText embedding model after adding MedTCS module on datasets of Table 1 for NER task.

the rankings by human experts to compute the Spearman (Sp) correlation coefficients with SciPy (Virtanen et al., 2020).

In extrinsic evaluation, the task of tagging the biomedical entities is performed by using a machine learning model trained on the encoded vectors. The performance measures used for this are recall, precision, and F1_score (Eqs 2–4).

3 Results

We compared the semantic and conceptual functionality of MedTCS with the pre-trained sub-word models (derivative models of BERT Devlin et al. (2019) and FastText Bojanowski et al. (2017)) for the biomedical and clinical domains.

3.1 Intrinsic evaluation

We evaluated the capability of MedTCS to enable the pre-trained word embedding models for encoding the OOV terms. MedTCS assisted the pre-trained word embedding models to achieve full coverage of all the conceptual pairs in the datasets. Moreover, we compared our model with related embedding models trained with FastText and BERT algorithms.

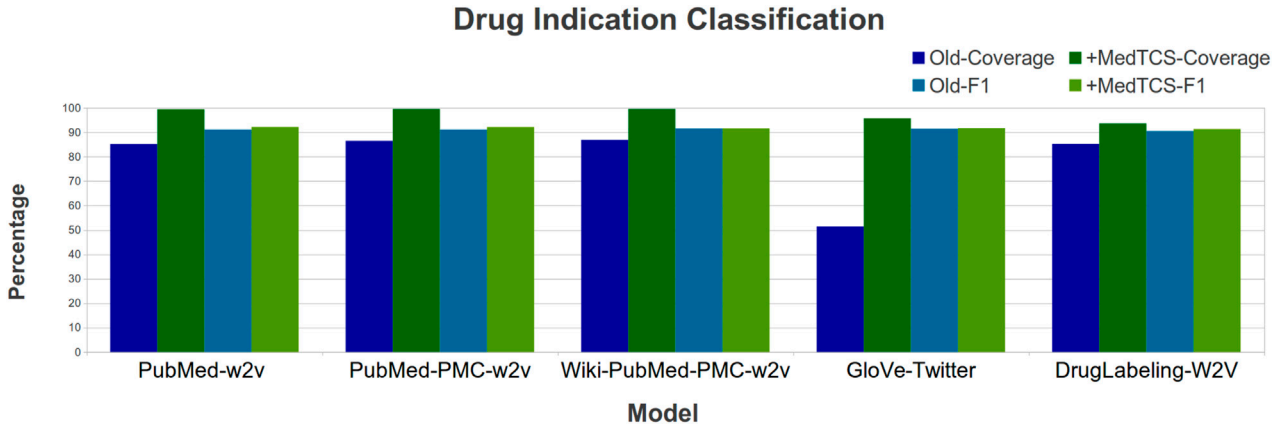


FIGURE 7
Model performances enhanced with MedTCS for Drug indication classification.

In this experiment, we included popular embedding models as a baseline, such as BioWordVec (Zhang et al., 2019), BioNLP (Chiu et al., 2016), PubMed-w2v, PubMed-PMC-w2v, and Wiki-PubMed-PMC-w2v (Moen and Ananiadou, 2013), most of which are defined under the Word2Vec algorithm (Mikolov et al., 2013b). Our baseline has the same encoder and decoder method for the NLP task without the MedTCS module. As the selected datasets include multi-word terms; therefore, the average vectors of each word are calculated with and without the MedTCS module.

Our analysis showed that the MedTCS module enabled all pre-trained embedding models to achieve full coverage with persuasive correlation scores on all datasets (Figure 3). For example, on the EHR-RelB dataset, the coverage of the *BiowordVec model* was enhanced from 2,857 terms pair to 3,630 terms pair and the Spearman (Sp) correlation also improved from 0.393 to 0.405. Overall, our results show that all models achieved 100% coverage of all the datasets with a slight decrease in correlation scores. As the OOV words are being approximated, therefore, a slight decrease in correlation scores is naturally expected.

We also enhanced the latest clinical word embedding models with MedTCS. The PMC Open Access Subset-Case reports (OA-CR) embedding models trained using word2vec/GloVe encountered the OOV word problem while working on the UMNSRS-Similarity dataset (Flamholz et al., 2022). MedTCS improved the coverage of all word embedding models from approximately 62% → 98% Supplementary Case S1.

Similarly, we analyzed the functionality of FastText to handle the OOV problem on the PMC Open Access subsets - Clinical Report (OA-CR) models and the PMC Open Access subsets - all manuscripts (OA-All) models (Flamholz et al., 2022). FastText trains each word vector along with its n-gram vectors. In the case of any OOV word, the average of its n-gram vectors are used to encode it (Bojanowski et al., 2017). For the FastText based OA-CR-600 embedding model, the Spearman (Sp) correlation value improved from 0.38 → 0.47 Supplementary Case S2. In conclusion, the MedTCS module enabled the different variants of OA-CR models to encode the vector for OOV terms from its search space effectively.

The OA-CR models have a small vocabulary; MedTCS enabled these models to achieve 100% coverage on all datasets as shown in Figure 4. Moreover, the MedTCS assisted the OA-CR models and the OA-ALL models to have significantly improved correlation scores, e. g., the FastText OA-All-300d model on the EHR-RelB dataset achieved 100% coverage and improved the Spearman (Sp) correlation scores from 0.25 → 0.35. The results on other variants of the OA-CR and OA-ALL embedding models for intrinsic evaluation are similar, as shown in Supplementary Figure S3.

On the other hand, BERT models use sub-words to solve the OOV word problem. We compared the proposed model with BERT and its derivative models defined for the clinical and biomedical domain [available on HuggingFace Wolf et al. (2019); Wolf et al. (2020)]. MedTCS outperformed BERT-based models by a significant margin in terms of correlation scores on the UMNSRS-Similarity dataset (Table 2). Moreover, in (Table 3), we have compared our best achieved results with recently reported scores of UMNSRS datasets (Mao and Fung, 2020; Singh and Jin, 2020; Yuan et al., 2022). MedTCS achieved significantly better coverage and correlation scores.

3.2 Extrinsic evaluation

Extrinsic evaluation requires training a system for the related downstream NLP tasks like NER, classification, etc,. The existing word embedding models achieve sub-optimal results due to the ineffective handling of OOV words (encoded unknown words with their n-gram vectors or a randomly generated vector). We

TABLE 4 Examples of the sub-word tokenization schemes followed by the different algorithms with the medical terminology-based MedTCS module.

Term	MedTCS	FastText	BioBert	CODER
		Bojanowski et al. (2017)	Lee et al. (2020)	Yuan et al. (2022)
mastodynia	breast, pain discomfort	<ma,mas,ast sto,tod,ody dyn.yni,nia,ia>	[CLS],mast,## ody,##nia, [SEP]	[CLS],mast,## odynia, [SEP]
prostatism	prostate, gland state,of,or,condition	<pr,pro,ros,ost,sta tat,ati,tis,ism,sm>	[CLS],pro,##sta ##tism, [SEP]	[CLS],prost,## atism, [SEP]
prostatorrhea	prostate, gland flow, excessive discharge	<pr,pro,ros,ost.sta tat,ato,tor,orr,rrh rhe,hea,ea>	[CLS],pro,##sta ##tor,##r,##hea [SEP]	[CLS],prost,## ator,##rh,##ea [SEP]
blepharospasm	eyelid,or,eyelash sudden,or involuntary	<bl,ble,lep,eph,pha har,aro,ros,osp,spa pas,asm,asm>	[CLS],b,##le,## pha,##ros,## pas,##m, [SEP]	[CLS],ble,## pha,#rosp,## asm, [SEP]
dyslipidemia	painful,fat,a blood, condition	<dy,dys,ysl,sli,lip pii,pid,ide,dem emi,mia,ia>	[CLS],d,##ys,## lip,##ide,## mia, [SEP]	[CLS] dyslipidemia [SEP]
dyspnea	painful, breathing	<dy,dys,ysp,spn pne,nea,ea>	[CLS],d,##ys,## p,##nea, [SEP]	[CLS],dyspnea [SEP]
urethrorrhea	urethra, flow excessive discharge	<ur,ure,ret,eth,thr hro,ror,orr,rrh,rhe hea,ea>	[CLS],u,##ret,## hr,##or,##r,## hea, [SEP]	[CLS],ureth,## ro,##r,##rh,## ea, [SEP]
arteriosclerosis	artery, hardening	<ar,art,rte,ter,eri rio,ios,osc,scl,cle ler,ero,ros,osi,sis,is>	[CLS],art,##eri ##os,##cle,## rosis, [SEP]	[CLS],arterio ##sc,##ler,## osis, [SEP]
dermatitis	Skin inflammation	<de,der,erm,rma mat,ati,tit,its,ts>	[CLS],der,##mat ##itis, [SEP]	[CLS] dermatitis [SEP]

tested the enriched vectors (by MedTCS) in identifying disease names from documents. We trained a bidirectional long-short term memory with a convolutional neural network (BiLSTM-CNN) Chiu and Nichols (2016) on the annotated corpus of BC5CDR and NCBI-disease (Table 1).

Figure 5 showed the performance enhancement in terms of coverage and F1 score (in percentage) achieved after replacing the randomly generated vector approach with our MedTCS module for OOV words. The MedTCS module with PubMed-w2v embedding enabled improved the coverage up to 13% on the NCBI-Disease dataset. Overall, MedTCS enabled word embedding models to achieve 100% coverage with an improved F1 score.

Similarly, we compared the MedTCS module with the n-gram approach for the NER task. Figure 6 showed that MedTCS improved the F1-score between 10 and 20% for the various embedding models as compared to the FastText n-gram vectors under the same parameters as for the BiLSTM NER system (Chiu and Nichols, 2016). The FastText OA-All-300d model with MedTCS achieved an F1-score of 0.80 (an improvement of 0.35) on the BC5CDR corpus and an F1-score of 0.81 (an improvement of 0.32) for the NCBI-disease

corpus. Similar results were achieved on the other variants of the OA-CR and OA-ALL embedding models for the NER task (Supplementary Figure S4).

Bhatt et al. (2021) recently developed a Drug Indication Classification and Encyclopedia (DICE) based on FDA approved human prescription drug labeling. They also generated "DrugLabelling-W2V" embeddings based on Word2Vec and used them to classify each sentence into one of the five classes (indications, contradictions, side effects, usage instructions, and clinical observations). We enhanced the "DrugLabelling-W2V" embedding with the MedTCS module and improved the coverage by 9% and the F1_score by 1% (Figure 7).

4 Discussion

Curating a large corpus is the traditional approach in NLP to cover more concepts and enhance the vocabulary of word-level embedding models. For example, meta-data from dictionaries, meta-thesaurus, and hierarchical relationships from ontologies were also used as corpus. In the biomedical and clinical domains, the larger corpus of PubMed-PMC from MEDLINE and Wikipedia (Denoyer and Gallinari, 2006) was used to enlarge the vocabulary. Similarly, the NCBI sources, including the Medical Subject Heading (MeSH) (Lipscomb, 2000), the Unified Medical Language System (UMLS) metathesaurus concepts (NLM, 2004), and the Systemized Nomenclature of Medicine—Clinical Terms (SNOMED CT) concepts (Donnelly, 2006) have also been used as meta-corpus. The semantic content of the ontologies and the meta-data like Web Ontology Language (OWL) has also been used to train embedding vectors (Grau et al., 2008). In spite of these efforts, while encoding some rare terms and concepts, the embedding models still faced the OOV problem like in the BioWordVec embedding model (Zhang et al., 2019).

We have developed MedTCS, a module that generates the vector representation for unknown words based on medical knowledge. Different approximation techniques derived from medical knowledge bases have been used to encode the OOV words. To the best of our knowledge, this is the first-ever post-processing and run-time solution for the OOV problem that is specifically designed for pre-trained biomedical/clinical word embedding models. Each OOV word is parsed into its components, which are replaced with their meanings to generate the semantic vectors. In addition, MedTCS's segmentation model tokenizes compound words into its word units, as shown in Figure 2. The MedTCS module outperforms the FastText n-gram approach to handle OOV words as shown in Figure 4.

In an empirical analysis of the BERT and its derivative models, we have observed that these models can have a high cosine similarity value between pairs given in datasets (Table 1). However, in the task to measure the degree of contextual relatedness and similarity between biomedical and clinical terms, they showed decreased performance (Table 2). Furthermore, according to our findings on BERT models, CODER Yuan et al. (2022) has better performance, probably because it encodes most of the words without splitting them into their sub-words, as shown in Table 4.

Word embedding models are of great importance for various biomedical NLP applications, however they currently face a major problem of assigning vectors for unknown and rare words. To fill this gap, we have developed the MedTCS module to facilitate the pre-trained word representation models in encoding medical terms. We hope that our module will be considered as a standard medical term tokenizer for the application of NLP in the biomedical domain. MedTCS can also allow other biomedical NLP researchers to develop knowledge-based modules in a variety of real-world applications. Moreover, our research highlighted that there is a need to not only train large embedding models but also some knowledge-driven modules for the medical and clinical domains. According to our knowledge, MedTCS is the first post-processing and run-time solution for the OOV problem that improves the applicability and semantic efficiency of pre-trained embedding of medical terms.

Data availability statement

The code for our module and instructions for the use can be found on GitHub, https://github.com/NadiaSaeed/MedTCS.git. Publicly available datasets used for performance evaluation and analysis during the current study are available at https://github.com/cambridgeltl/MTL-Bioinformatics-2016. The original contributions presented in the study are publicly available. The generated dictionaries of MedTCS are available on GitHub, https://github.com/NadiaSaeed/MedTCS.git.

Author contributions

NS proposed and designed the study. NS and HN performed data analysis. NS and HN wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding

This research work was funded by the Higher Education Commission of Pakistan and the Ministry of Planning

Development and Reforms under the umbrella of the National Center in Big Data and Cloud Computing (NCBC).

Acknowledgments

The authors acknowledge the support of Amna Binte Kamran, Hassan Mustafa, and Umair Ayub in proof reading the manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.928530/full#supplementary-material

References

Alsentzer, E., Murphy, J. R., Boag, W., Weng, W.-H., Jin, D., Naumann, T., et al. (2019). "Publicly available clinical BERT embeddings," in Proceedings of the 2nd clinical natural language processing workshop (Minneapolis, MN: Association for Computational Linguistics), 72–78.

Banay, G. L. (1948). An introduction to medical terminology, Greek and Latin derivations. *Bull. Med. Libr. Assoc.* 36, 1–27.

Beltagy, I., Lo, K., and Cohan, A. (2019). "Scibert: a pretrained Language Model for scientific text," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing EMNLP, Hong Kong, China, November 3–7, 2019 (Association for Computational Linguistics)

Bhatt, A., Roberts, R., Chen, X., Li, T., Connor, S., Hatim, Q., et al. (2021). Dice: a drug indication classification and encyclopedia for ai-based indication extraction. *Front. Artif. Intell.* 4, 711467. doi:10.3389/frai.2021.711467

Bird, S., and Loper, E. (2004). *Nltk: the natural language toolkit*. Barcelona, Spain: Association for Computational Linguistics.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.* 5, 135–146. doi:10.1162/tacl_a_00051

Boukkouri, H. E., Ferret, O., Lavergne, T., Noji, H., Zweigenbaum, P., and Tsujii, J. (2020). "CharacterBERT: Reconciling ELMo and BERT for word-level open-vocabulary representations from characters," in Proceedings of the 28th international conference on computational linguistics (Barcelona, Spain: International Committee on Computational Linguistics.), 6903–6915.

Chiu, J. P., and Nichols, E. (2016). Named entity recognition with bidirectional lstm-cnns. *Trans. Assoc. Comput. Linguist.* 4, 357–370. doi:10.1162/tacl_a_00104

Chiu, B., Crichton, G., Korhonen, A., and Pyysalo, S. (2016). "How to train good word embeddings for biomedical nlp," in Proceedings of the 15th Workshop on Biomedical Natural Language Processing, Berlin, Germany, August 7–12, 2016, 166–174.

Cohen, K. B., and Demner-Fushman, D. (2014). *Biomedical natural language processing*, 11. Amsterdam: John Benjamins Publishing Company.

Denoyer, L., and Gallinari, P. (2006). "The wikipedia xml corpus," in International Workshop of the Initiative for the Evaluation of XML Retrieval, New York, NY, June 1, 2006 (Springer), 12–19.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). "BERT: Pre-training of deep bidirectional transformers forlanguage understanding," in Proceedings of NAACL-HLT 2019, June 2–7, 2019 (Minneapolis, MN: Association for Computational Linguistics), 4171–4186.

Donnelly, K. (2006). Snomed-ct: the advanced terminology and coding system for ehealth. *Stud. Health Technol. Inf.* 121, 279–290.

Flamholz, Z. N., Crane-Droesch, A., Ungar, L. H., and Weissman, G. E. (2022). Word embeddings trained on published case reports are lightweight, effective for clinical tasks, and free of protected health information. *J. Biomed. Inf.* 125, 103971. doi:10.1016/j.jbi.2021.103971

Grau, B. C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U., et al. (2008). Owl 2: the next step for owl. *J. Web Semant.* 6, 309–322. doi:10.1016/j.websem.2008.05.001

Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., et al. (2022). Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans. Comput. Health* 3 (1), 1–23. doi:10.1145/3458754

Henderson, B., and Dorsey, J. L. (2019). *Medical terminology for dummies*. Hoboken: John Wiley & Sons.

Huang, K., Altosaar, J., and Ranganath, R. (2019). Clinicalbert: modeling clinical notes and predicting hospital readmission. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/1904.05342.

Jin, Q., Dhingra, B., Cohen, W. W., and Lu, X. (2019). "Probing biomedical embeddings from language models," in Proceedings of the 3rd workshop on evaluating vector space representations for NLP (Minneapolis, MN: Association for Computational Linguistics), 82–89.

Kalyan, K. S., and Sangeetha, S. (2021). A hybrid approach to measure semantic relatedness in biomedical concepts. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/2101.10196.

Leaman, R., Khare, R., and Lu, Z. (2015). Challenges in clinical natural language processing for automated disorder normalization. *J. Biomed. Inf.* 57, 28–37. doi:10.1016/j.jbi.2015.07.010

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2020). Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 1234–1240. doi:10.1093/bioinformatics/btz682

Lipscomb, C. E. (2000). Medical subject headings (mesh). *Bull. Med. Libr. Assoc.* 88, 265–266.

Mao, Y., and Fung, K. W. (2020). Use of word and graph embedding to measure semantic relatedness between unified medical language system concepts. *J. Am. Med. Inf. Assoc.* 27, 1538–1546. doi:10.1093/jamia/ocaa136

MedicineNet (2007). [Dataset]. Medical termsmedical dictionary, medical definitions medical terminology. Available at: https://www.medicinenet.com (Accessed 01 13.2022).

Merriam-Webster (2018). [Dataset]. Dictionary by merriam-webster America's most-trusted online dictionary. Available at: https://www.merriam-webster.com (Accessed 01 13.2022).

Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., and Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb. Med. Inf.* 17, 128–144. doi:10.1055/s-0038-1638592

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Adv. neural Inf. Process. Syst.* 26, 1–9. doi:10.48550/arXiv.1310.4546

Moen, S., and Ananiadou, T. S. S. (2013). "Distributional semantics resources for biomedical text processing," in Proceedings of LBM, Tokyo, Japan, December 12–13, 2013, 39–44.

NLM (2004). *Unified medical language system, rxnorm*. New Jersey, United States: National Library of Medicine. Available at:.

Pakhomov, S., McInnes, B., Adam, T., Liu, Y., Pedersen, T., and Melton, G. B. (2010). Semantic similarity and relatedness between clinical terms: an experimental study. *AMIA Annu. Symp. Proc.* 2010, 572. American Medical Informatics Association. doi:10.13020/D6CX04

Pakhomov, S. V., Pedersen, T., McInnes, B., Melton, G. B., Ruggieri, A., Chute, C. G., et al. (2011). Towards a framework for developing semantic relatedness reference standards. *J. Biomed. Inf.* 44, 251–265. doi:10.1016/j.jbi.2010.10.004

Peng, Y., Yan, S., and Lu, Z. (2019). . "Transfer learning in biomedical natural language processing: an evaluation of bert and elmo on ten benchmarking datasets," in Proceedings of the 18th BioNLP workshop and shared task (Florence, Italy: Association for Computational Linguistics), 58–65.

Pennington, J., Socher, R., and Manning, C. D. (2014). "Glove: global vectors for word representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, October 25–29, 2014 (EMNLP), 1532–1543.

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., et al. (2018). . "Deep contextualized word representations," in Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies (New Orleans, LA: Association for Computational Linguistics), 2227–2237.

Řehřek, R., and Sojka, P. (2011). *Gensim—statistical semantics in python*. Paris: EuroScipy. *Retrieved from genism. org.*

Schulz, C., Levy-Kramer, J., Van Assel, C., Kepes, M., and Hammerla, N. (2020). . "Biomedical concept relatedness–a large ehr-based benchmark," in Proceedings of the 28th international conference on computational linguistics, December 8–13, 2020 (Barcelona, Spain), 6565–6575.

Singh, A., and Jin, W. (2020). "On using composite word embeddings to improve biomedical term similarity," in 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), October 26–28, 2020 (IEEE), 281–287.

Smit, P., Virpioja, S., Grönroos, S.-A., and Kurimo, M. (2014). "Morfessor 2.0: toolkit for statistical morphological segmentation," in The 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Gothenburg, Sweden, April 26-30, 2014 (Aalto University).

Subramanyam, K. K., and Sangeetha, S. (2020). Deep contextualized medical concept normalization in social media text. *Procedia Comput. Sci.* 171, 1353–1362. doi:10.1016/j.procs.2020.04.145

Virpioja, S., Smit, P., Grönroos, S.-A., and Kurimo, M. (2013). *Morfessor 2.0: python implementation and extensions for morfessor baseline*. Gothenburg, Sweden: Association for Computational Linguistics.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2

Wang, J., Liang, S.-l., and Ge, G.-c. (2008). Establishment of a medical academic word list. *Engl. Specif. Purp.* 27, 442–458. doi:10.1016/j.esp.2008.05.003

Wang, X., Zhang, Y., Ren, X., Zhang, Y., Zitnik, M., Shang, J., et al. (2019). Cross-type biomedical named entity recognition with deep multi-task learning. *Bioinformatics* 35, 1745–1752. doi:10.1093/bioinformatics/bty869

WebMD (2012). [Dataset]. Webmd better information. better health. Available at: https://www.webmd.com (Accessed 01 13.2022).

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., et al. (2019). Huggingfaceś transformers: state-of-the-art natural language processing. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/1910.03771.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., et al. (2020). "Transformers: state-of-the-art natural language processing," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, November 16–20, 2020. 38–45.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., et al. (2016). Google's neural machine translation system: bridging the gap between human and machine translation. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/1609.08144.

Yuan, Z., Zhao, Z., Sun, H., Li, J., Wang, F., Yu, S., et al. (2022). Coder: knowledge-infused cross-lingual medical term embedding for term normalization. *J. Biomed. Inf.* 126, 103983. doi:10.1016/j.jbi.2021.103983

Zhang, Y., Chen, Q., Yang, Z., Lin, H., and Lu, Z. (2019). Biowordvec, improving biomedical word embeddings with subword information and mesh. *Sci. Data* 6, 52. doi:10.1038/s41597-019-0055-0

Zhu, H., Paschalidis, I. C., and Tahmasebi, A. (2018). Clinical concept extraction with contextual word embedding. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/1810.10566.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 17 August 2022
DOI 10.3389/fgene.2022.958092

Check for updates

OPEN ACCESS

EDITED BY
Wei Wang,
Huzhou Maternity & Child Health Care Hospital, China

REVIEWED BY
Jun Xu,
The First Affiliated Hospital of University of Science and Technology of China
Anhui Provincial Hospital, China
Min-Er Zhong,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

*CORRESPONDENCE
Changdong Sun,
sunchang19770110@163.com
Yonggui Han,
hanyonggui5968@163.com

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

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 31 May 2022
ACCEPTED 19 July 2022
PUBLISHED 17 August 2022

CITATION
Chen L, Gu H, Zhou L, Wu J, Sun C and Han Y (2022), Integrating cell cycle score for precise risk stratification in ovarian cancer.
Front. Genet. 13:958092.
doi: 10.3389/fgene.2022.958092

COPYRIGHT
© 2022 Chen, Gu, Zhou, Wu, Sun and Han. 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.

Integrating cell cycle score for precise risk stratification in ovarian cancer

Lingying Chen[1†], Haiyan Gu[1†], Lei Zhou[1], Jingna Wu[1], Changdong Sun[1*] and Yonggui Han[2*]

[1]Department of Obstetrics and Gynecology, Beilun District People's Hospital, Ningbo, China, [2]Department of Obstetrics and Gynecology, Beilun No 3 People's Hospital, Ningbo, China

Background: Ovarian cancer (OC) is a highly heterogeneous disease, of which the mesenchymal subtype has the worst prognosis, is the most aggressive, and has the highest drug resistance. The cell cycle pathway plays a vital role in ovarian cancer development and progression. We aimed to screen the key cell cycle genes that regulated the mesenchymal subtype and construct a robust signature for ovarian cancer risk stratification.

Methods: Network inference was conducted by integrating the differentially expressed cell cycle signature genes and target genes between the mesenchymal and non-mesenchymal subtypes of ovarian cancer and identifying the dominant cell cycle signature genes.

Results: Network analysis revealed that two cell cycle signature genes (*POLA2* and *KIF20B*) predominantly regulated the mesenchymal modalities of OC and used to construct a prognostic model, termed the Cell Cycle Prognostic Signature of Ovarian Cancer (CCPOC). The CCPOC-high patients showed an unfavorable prognosis in the GSE26712 cohort, consistent with the results in the seven public validation cohorts and one independent internal cohort (BL-OC cohort, qRT-PCR, n = 51). Functional analysis, drug-sensitive analysis, and survival analysis showed that CCPOC-low patients were related to strengthened tumor immunogenicity and sensitive to the anti-PD-1/PD-L1 response rate in pan-cancer (r = −0.47, OC excluded), which indicated that CCPOC-low patients may be more sensitive to anti-PD-1/PD-L1.

Conclusion: We constructed and validated a subtype-specific, cell cycle-based prognostic signature for ovarian cancer, which has great potential for predicting the response of anti-PD-1/PD-L1.

KEYWORDS

ovarian cancer, cell cycle, mesenchymal subtype, prognosis, PD-1/PD-L1

Introduction

Ovarian cancer (OC) is the leading cause of cancer death in women (Siegel et al., 2019). Due to lack of effective early screening methods and lack of obvious symptoms, most of the patients were diagnosed at an advanced stage, resulting in an overall 5-year survival rate of less than 50% (Lheureux et al., 2019). Clinical risk assessment factors include tumor stage, tumor grade, histopathological classification, de-bulking status, etc. Despite the good initial treatment effect, most ovarian cancer patients still suffer from tumor recurrence and eventually develop drug resistance to chemotherapy (Coleman et al., 2013). Currently, serum CA-125 level is a clinical biomarker for risk assessment of ovarian cancer. Due to its low specificity, the overall assessment effect is not as expected (Bottoni and Scatena, 2015). The high degree of heterogeneity and aggressiveness of OC often leads to treatment failure (Cancer Genome Atlas Research Network, 2011; Konecny et al., 2014). Therefore, there is a need to integrate tumor heterogeneity to identify novel prognostic predictors for OC.

Gene expression-based biomarkers for cancer risk assessment have been extensively explored (Mo et al., 2020). Several studies have established OC prognostic biomarkers based on gene expression (Pan and Ma, 2020; Yang et al., 2021). However, due to the heterogeneity of OC, most of the biomarkers have low prognostic efficacy and cannot be directly used in clinical practice. Recently, four ovarian cancer molecular subtypes with distinct molecular and clinical characteristics were found (Cancer Genome Atlas Research Network, 2011), among which, the mesenchymal subtype had the poorest prognosis. Afterward, the mesenchymal subtype is consistent in several other subtyping systems (Konecny et al., 2014; Chen et al., 2018). Importantly, the mesenchymal subtype of OC shows poor clinical outcomes, indicating the need to integrate the intrinsic modalities of this malignant subtype for risk management in OC.

Cancer manifests itself as an infinite proliferation of cells, the main reason for which is related to improper cell cycle regulation (Williams and Stoeber, 2012). The cell cycle is precisely regulated by cyclin-dependent kinases (CDKs) (Bertoli et al., 2013). However, relevant cell cycle-based biomarkers are rare and still lacking in ovarian cancer. Considering the highly heterogeneous nature of OC, by integrating mesenchymal modalities and the cell cycle signature underlying the mesenchymal subtype, a network-based approach was adopted to identify the dominant cell cycle signature, which regulates the most aggressive OC subtype. Subsequently, we established a prognostic model, termed Cell Cycle Prognostic Signature of Ovarian Cancer (CCPOC), and exploration of the prognosis capacity of CCPOC in OC. Our signature incorporates cell cycle system and tumor heterogeneity and would be used to screen OC patients who may benefit from a more precise treatment.

Materials and methods

Public dataset preparation and preprocessing

We obtained 1,798 OC samples from eight publicly available datasets. The training dataset was the GSE26712 (Bonome et al., 2008) cohort ($n = 182$). Validation cohorts were the TCGA(Cancer Genome Atlas Research Network, 2011) ($n = 578$), GSE9891 (Tothill et al., 2008) ($n = 285$), ICGC-AU (Patch et al., 2015) ($n = 111$), GSE138866 (Hu et al., 2020) ($n = 130$), GSE32062 (Yoshihara et al., 2012) ($n = 260$), GSE14764 (Denkert et al., 2009) ($n = 80$), and GSE51088 (Karlan et al., 2014) ($n = 172$) datasets. Together with the corresponding clinical information, the normalized expression datasets sourced from the GEO database were downloaded *via* the GEOquery package (version 2.58.0). The transcription data (Affymetrix U133A) and relevant clinical information on TCGA were retrieved from the Firebrowse (http://firebrowse.org/) database. The standardized expression profile and clinical information of ICGC-AU were downloaded from the International Cancer Genome Consortium (ICGC, https://icgc.org) OV-AU (Ovarian cancer-Australia) database. For the microarray data, the gene expression data probe IDs were transformed into gene symbols; if multiple probe IDs were mapped to the same gene symbol, the one with the highest average value was selected. The molecular subtyping information was retrieved from Verhaak's study (Verhaak et al., 2010). The detailed clinical parameters of all cohorts are listed in Table 1.

Clinical samples

For the independent internal validation cohort (BL-OC cohort), we retrospectively collected 51 formalin-fixed paraffin-embedded (FFPE) blocks from patients who underwent surgery in Beilun People's Hospital (from 1st January, 2015 to 1st January, 2021), Ningbo, China. Criteria for patient sample selection: longer follow-up (> 5 years) and had evaluation of adjuvant chemotherapy efficacy and no history of cancer other than ovarian cancer. This study was approved by the Ethics Committee of the Beilun People's Hospital.

Network analysis screening key regulated cell cycle genes for the mesenchymal subtype

We obtained 313 cell cycle-related genes (CRGs) through the concatenation of the cycle-related genes from the MSigDB database (Version 7.2; KEGG cell cycle pathway, HALLMARK G2M pathway) and Cuzick's study (Cuzick et al., 2011). CRGs with expression in all datasets were retained for subsequent analysis. We integrated differentially expressed target genes and cell cycle

TABLE 1 Overview of the clinical and pathologic characteristics of all the datasets.

	Training cohort	Public validation cohorts							Internal validation
	GSE26712	TCGA	GSE9891	ICGC-AU	GSE138866	GSE32062	GSE14764	GSE51088	BL-OC
Number of patients	182	578	285	111	130	260	80	172	51
Age (years)									
Mean, years (STD)	62 (11.9)	60 (11.6)	60 (10.6)	59 (8.7)	62 (11.9)			58 (12.6)	58 (11.2)
Histopathology									
Serous	182 (100%)	568 (98%)	264 (93%)	111 (100%)	130 (100%)	260 (100%)	68 (85%)	122 (71%)	58 (100%)
Others		10 (2%)	21 (7%)				12 (15%)	50 (29%)	
Stage									
I		16 (3%)	24 (8%)				8 (10%)	22 (13%)	
II		27 (5%)	18 (6%)		2 (2%)		1 (1%)	9 (5%)	
III	144 (79%)	436 (75)	218 (76%)	96 (86%)	106 (82%)	204 (78%)	69 (86%)	103 (60%)	
IV	36 (20%)	84 (15%)	22 (8%)	15 (14%)	22 (17%)	56 (22%)	2 (3%)	17 (10%)	
Unknown	2 (1%)	15 (3%)	3 (1%)						
Grade									
Well		6 (1%)	19 (7%)				3 (4%)	16 (9%)	
Moderately		69 (12%)	97 (34%)	15 (14%)		131 (50%)	23 (29%)	14 (8%)	
Poorly		479 (83%)	163 (57%)	66 (59%)	130 (100%)	129 (50%)	54 (68%)	119 (69%)	
Unknown		23 (4%)	6 (2%)	30 (27%)				23 (13%)	
Debulking									
Optimal	88 (48%)	367 (63%)	160 (56%)		107 (82%)	103 (40%)	39 (49%)		
Suboptimal	94 (52%)	140 (24%)	88 (31%)		15 (12%)	157 (60%)	23 (29%)		
Unknown		71 (12%)	37 (13%)		8 (6%)		18 (22%)		
Vital status									
Alive	55 (30%)	270 (47%)	169 (59%)	23 (21%)	31 (24%)	139 (53%)	59 (74%)	40 (23%)	16 (31%)
Dead	127 (70%)	290 (50%)	113 (40%)	88 (79%)	99 (76%)	121 (47%)	21 (26%)	112 (65%)	35 (69%)
NA		18 (3%)	3 (1%)					20 (12%)	
Median OS, months (±SE)	38.2 (2.6)	29.4 (1.1)	28.5 (1.4)	32.4 (3.0)	33.7 (4.6)	41.5 (1.5)	35.0 (1.7)	49.7 (4.0)	37.5 (4.6)

genes between the mesenchymal subtype and other subtypes and performed a network analysis by using the RTN package (version 2.10.0) to infer and investigate the relationship between cell cycle genes and potential target genes. Specifically, the network analysis consists of three parts: first, the mutual information (MI) between a cell cycle signature gene and all potential target genes is calculated, and insignificant associations are removed by permutation analysis; second, unstable interactions are removed by bootstrapping; and finally, the ARACNe algorithm is applied to reduce redundant indirect regulations. Together, the GSE26712 dataset was used as the training cohort. Univariate Cox regression analysis screened 34 cell cycle genes (p <0.1) for a subsequent analysis. Subsequently, 12 cell cycle genes (|log2 FC| > 0.25, BH-adjusted p < 0.05) and 1,704 target genes (log2 FC > 0.25, BH-adjusted p < 0.05) were determined differentially expressed in the mesenchymal subtype compared with non-mesenchymal subtypes. Then, a master regulator analysis (Fletcher et al., 2013) (MRA) was performed to examine the overrepresentation of the mesenchymal signature in the regulation of each cell cycle gene by a hypergeometric test. After the hypergeometric test results for all cell cycle signature genes, adjusted p-values were calculated using the Benjamini–Hochberg procedure. Two cell cycle signature genes of top significance (Benjamini–Hochberg-adjusted p-value < 0.05) were selected as master regulators. For detailed calculation steps and calculation code, please refer to the Vignettes of the RTN package (bioconductor.org/packages/release/bioc/vignettes/RTN/inst/doc/RTN.html).

Development and evaluation of the risk model for ovarian cancer in public cohorts

Network analysis revealed that two cell cycle genes (*POLA2* and *KIF20B*) were the key regulators of the mesenchymal

subtype, which is the most aggressive subtype of ovarian cancer. Subsequently, the multivariable Cox regression model was used to construct a prognostic signature, termed Cell Cycle Prognostic Signature of Ovarian Cancer (CCPOC), in the GSE26172 cohort with these two signature genes. The risk score formula was constructed based on a linear combination of the expression levels weighted with the regression coefficients: CCPOC = (−0.6527 × *POLA2*) + (0.4975 × *KIF20B*). Based on the upper quantile score of each cohort calculated by the risk score formula, patients were divided into CCPOC-high and CCPOC-low subgroups. The prognostic relevance of CCPOC was evaluated in seven public independent validation datasets (TCGA, GSE9891, ICGC-AU, GSE138866, GSE32062, GSE14764, and GSE51088) with the Kaplan–Meier analysis. Univariate and multivariate analyses were performed with other clinical factors to test whether the CCPOC can be considered an independent prognostic predictor.

Validation of the signature genes in the internal ovarian cancer cohort by quantitative reverse transcription PCR (qRT-PCR)

Fifty-one OC tissues were obtained from Beilun People's Hospital. This study was approved by the Ethics Committee of Beilun People's Hospital. Total RNA was extracted by using the High Pure RNA paraffin kit (Roche Applied Science, Indianapolis, IN) from FFPE tissues of the BL-OC cohort. Reverse transcription was performed with High Capacity cDNA (Thermo Scientific). qRT-PCR was performed with the QuantStudio™ 12 K Flex Real-Time PCR System (Thermo Scientific) according to the manufacturer's recommended operating conditions. β-Actin was tested for data normalization. The primers of each gene are listed as follows: POLA2: F CACCACATC TGACAGCATCACG, R CCACCTGTTCATGCTTAGCAT CC; KIF20B: F GCTGACTTTAAGGAGACTCTGCT, R GTGGCACAAATGTCTTTCGCTGC; and β-Actin: F CAC CATTGGCAATGAGCGGTTC, R AGGTCTTTGCGGATG TCCACGT. The expression of each gene was calculated using the log2 (2−ΔΔCT) method.

Functional analysis

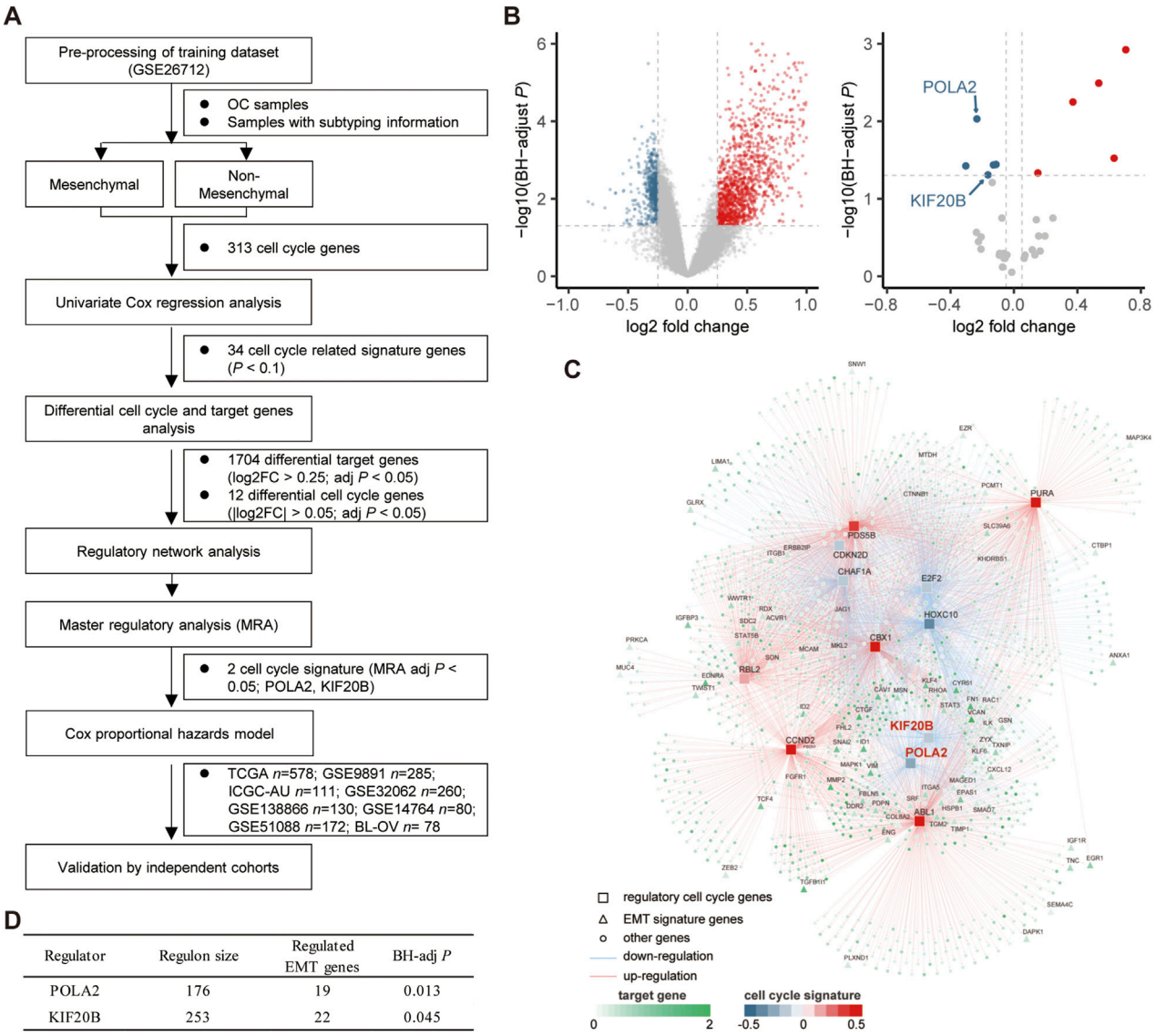
Gene Set Enrichment Analysis (GSEA) was carried out to test the dysregulated pathways in different CCPOC risk groups by using the HTSanalyzeR package (Wang et al., 2011) (version 2.3.5) with 1,000 permutations. Hallmark (h.all.v7.2.symbols.gmt) and KEGG (c2.cp.kegg.v7.2.symbols.gmt) gene sets were downloaded from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/). Only gene sets with >five genes were included in the analyses. To evaluate the immunobiological difference between different CCPOC risk groups, CIBERSORT (Newman et al., 2015), a de-convolution algorithm, was used to characterize 22 types of immune cell abundance for each sample. For the TCGA-OV mutation, data were downloaded from the cBioPortal database (https://www.cbioportal.org/).

Data sources for chemotherapy, immunotherapeutic, and pan-cancer analysis

GSE146965 (Jiménez-Sánchez et al., 2020) and PMID17290060 (Dressman et al., 2007) containing chemotherapy response information were used for chemotherapy sensitivity analysis. Two immunotherapeutic cohorts: the IMvigor210 cohort (Mariathasan et al., 2018) was an advanced urothelial cancer with the intervention of atezolizumab, an anti-PD-L1 antibody; the GSE78220 cohort (Hugo et al., 2016) was metastatic melanoma treated with pembrolizumab, an anti-PD-1 antibody. For the IMvigor210 cohort, expression data and clinical data were downloaded from https://github.com/SiYangming/IMvigor210CoreBiologies. The TCGA PanCancer Atlas gene expression profiles and clinical information were downloaded by the TCGAbiolinks package (version 2.18.0). The infiltration status of different immune cell populations in the TCGA PanCancer Atlas was downloaded from Tamborero's study (Tamborero et al., 2018). The marker genes of MHC, immunoinhibitory, and immunostimulatory molecules were reported by Charoentong et al. (2017). DNA damage response (DDR) signature genes were extracted from the study of Theo et al. (Knijnenburg et al., 2018). The expression of the proteins encoded by the signature genes was validated in the Human Protein Profiles (http://www.proteinatlas.org) database. The objective response rate (ORR) was obtained from public research (listed in Supplementary Table S6).

Statistical analysis

The immune genes and potential target genes between the mesenchymal and non-mesenchymal subtypes underwent differential analysis with the R limma package (version 3.42.2). Kaplan–Meier analysis was performed to test survival differences between different groups with the log-rank test using the R survival package (version 2.41.3). The prognostic value of the selected cell cycle signature was shown using the univariate Cox regression analysis. The independent prognostic effect of CCPOC was tested using univariate and



Regulator	Regulon size	Regulated EMT genes	BH-adj *P*
POLA2	176	19	0.013
KIF20B	253	22	0.045

FIGURE 1
Network inference identified two cycle signature genes (POLA2 and KIF20B) as the key regulators in the mesenchymal subtype of OC. **(A)** Study design of the present work. **(B)** Volcano plot of the differentially expressed target genes and cell cycle signature genes in the mesenchymal subtype. **(C)** Integrated network showing the relationships between the expression data of the cell cycle signature genes and target genes. **(D)** Master regulator analysis results.

multivariate Cox regression analyses. The survival prediction was assessed by the concordance indices (C-index) and the robust hazard ratio (D-index), which were calculated using the survcomp package. Student's t-tests and Kruskal–Wallis tests were used to conduct difference comparisons of three or more groups. The correlations between the CCPOC scores and the ORR were evaluated using Pearson's correlation. $p < 0.05$ was considered significant. All statistical analyses were performed in R (version 3.6.1, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

Results

The integrative analysis identifies two cell cycle genes as key regulators in the mesenchymal subtype

The mesenchymal subtype has the worst prognosis and shortest overall survival (Supplementary Figure S1). We intended to integrate the molecular modalities under this subtype to improve the OC risk assessment thereafter. Focusing on the mesenchymal subtype, we

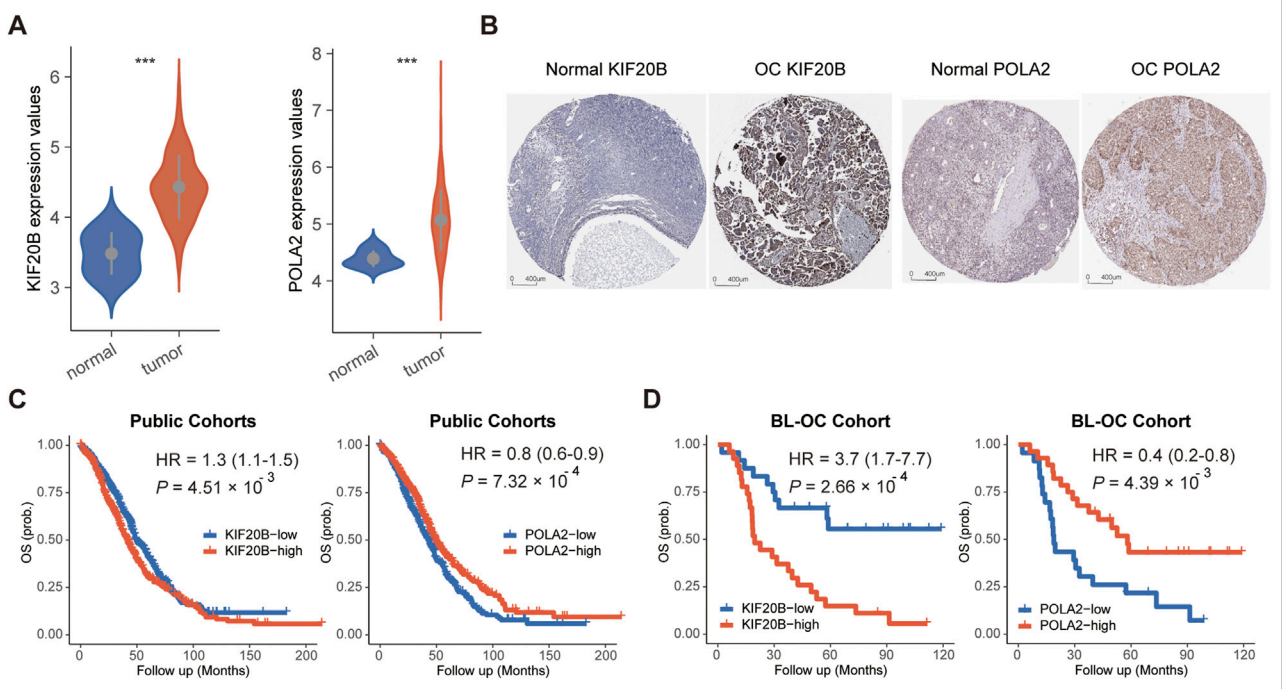


FIGURE 2
Expression and survival analyses for POLA2 and KIF20B in OC. **(A)** Expression levels of POLA2 and KIF20B in OC and normal tissues. **(B)** Protein levels encoded by POLA2 and KIF20B in normal and OC using clinical samples from the Human Protein Profiles. Survival analysis of POLA2 and KIF20B in public cohorts **(C)** and the BL-OC cohort **(D)**.

applied a network-based approach to investigate the regulatory role of the cell cycle, which is important in the progression of OC. Eight public datasets and one independent internal cohort with a total sample of 1,849 OC cases were included in this study (Table 1). 313 cell cycle-related genes (CRGs) (Supplementary Table S1) were downloaded from the MSigDB database and Cuzick's study (Cuzick et al., 2011). Based on the GSE26712 cohort, we performed an initial screening of the cell cycle genes by the univariate Cox regression analysis, and a total of 34 cell cycle genes (p <0.1) were screened for subsequent analysis. Subsequently, we conducted a differential analysis of the selected cell cycle genes and potential target genes between the mesenchymal and non-mesenchymal subtypes (Figure 1A). Twelve cell cycle genes (|log2 FC| > 0.25, BH-adjusted p < 0.05) and 1,704 target genes (log2 FC > 0.25, BH-adjusted p < 0.05) were determined to be differentially expressed in the mesenchymal subtype (Figure 1B). Based on the expression profiles of these prioritized cell cycle genes and target genes, we constructed a regulatory network by calculating the mutual information between a cell cycle gene signature and its potential targets (Figure 1C). Based on hypergeometric tests, a master regulator analysis (MRA) was performed to screen core regulators for the mesenchymal subtype (Supplementary Table S2). We identified 19 and 22 EMT genes enriched in the regulons of POLA2 (BH-adjusted p = 0.013) and KIF20B (BH-adjusted p = 0.045) (Figure 1D), respectively. Compared to the non-mesenchymal subtypes (immunoreactive, proliferative, and differentiated), the two candidate genes were significantly lower expressed in the mesenchymal subtype (Supplementary Figures S2A, B).

Compared to normal tissues, these two candidate cell cycle genes were all significantly highly expressed in OC tissues in the TCGA cohort (Figure 2A). Moreover, we checked the protein levels encoded by these two genes in the Human Protein Profiles database. POLA2 and KIF20B were moderately positive detected in OC clinical specimens when compared to their expression levels in normal samples (Figure 2B). Therefore, in the future, it is possible to evaluate the prognosis of OC patients by detecting the expression of these two genes on clinical specimens by IHC. Furthermore, the survival analysis revealed a prognostic association with overall survival in the public cohorts (Figure 2C) and the BL-OC cohort (Figure 2D). Together, the network-based approach identified two cell cycle genes, with a prognostic value, as key regulators in the mesenchymal subtype.

Construction and evaluation of the cell cycle prognostic signature in public cohorts and the BL-OC cohort

Based on the GSE26172 cohort, the risk model called "Cell Cycle Prognostic Signature of Ovarian Cancer" (CCPOC) was

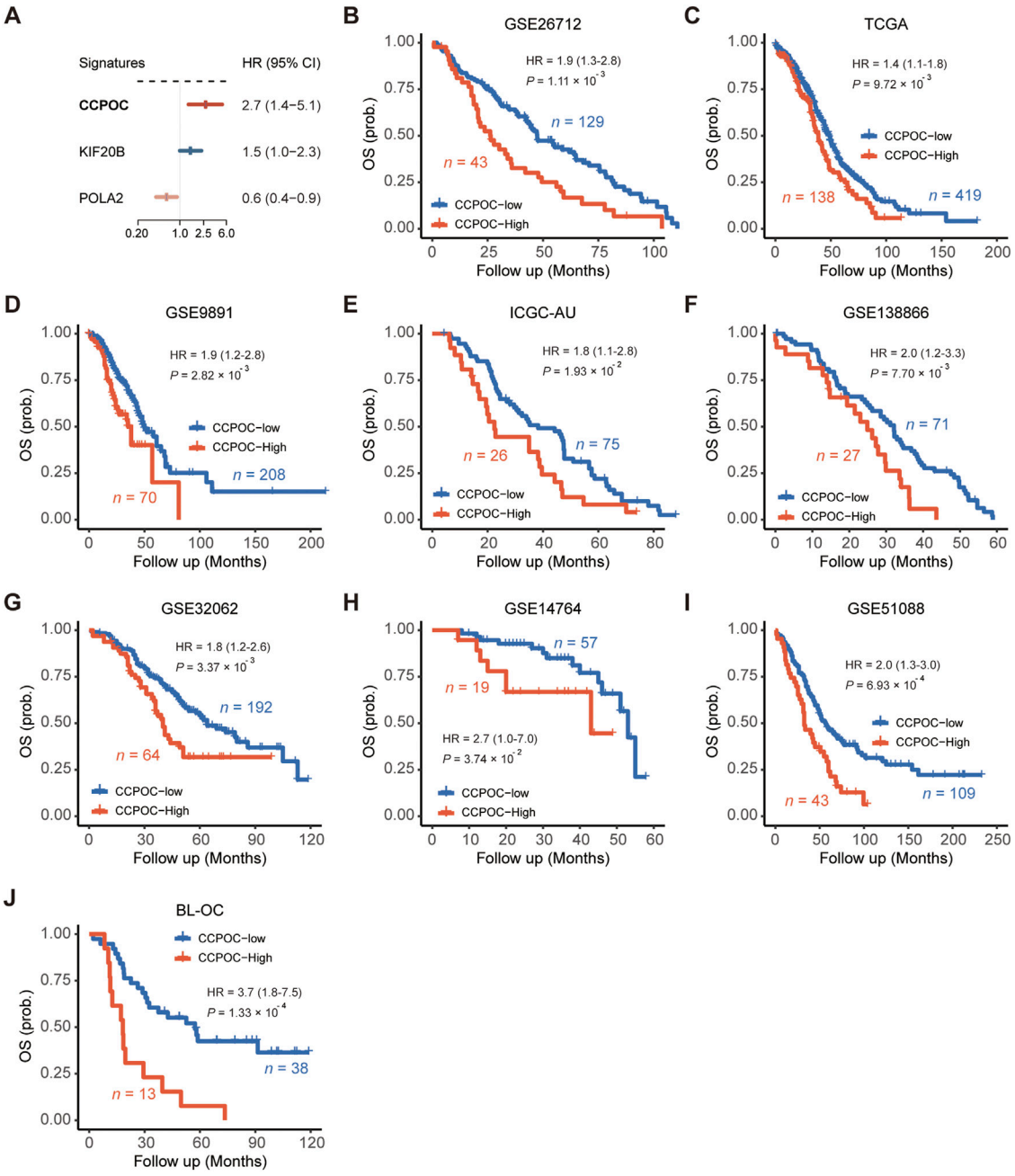


FIGURE 3
Assessment of the prognostic value of the CCPOC. **(A)** Comparison of prognostic efficiencies between the CCPOC and its individual constituents. **(B)** Kaplan–Meier survival analysis showing that the CCPOC-high group had an unfavorable OS in the training cohort (GSE26712). In the seven public validation cohorts **(C–I)**, the CCPOC-high group stably showed a significantly poor prognosis for OS. **(J)** Evaluation of the prognostic value of CCPOC in the BL-OC cohort.

constructed based on a linear combination of the expression levels weighted with the regression coefficients of these two cell cycle genes derived from the multivariate Cox regression analysis. Risk score = (−0.6527 × *POLA2*) + (0.4975 × *KIF20B*). Subsequently, risk scores were calculated for all patients in the public cohorts and our in-house validation BL-OC cohort (Supplementary Tables S3–S4). The CCPOC score showed prognostic efficiency with an AUC of 0.77 at 2 years and 0.79 at 5 years in the BL-OC cohort (Supplementary Figure S3). Based on the upper quantile score of each cohort calculated by

TABLE 2 Univariate and multivariate prognostic analyses of the cell cycle signature and clinicopathological factors in the training cohort and meta-validation of public cohorts.

	GSE26712				Meta-validation of public cohorts			
	Univariate		Multivariate		Univariate		Multivariate	
	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*
Grade* (3 vs. 1&2)	1.75 (0.44–7.07)	0.43	1.63 (0.40–6.58)	0.49	2.63 (1.52–4.55)	**5.0E-04**	2.09 (0.87–5.05)	0.10
Debulking (optimal vs. suboptimal)	1.32 (1.02–1.71)	**0.03**	1.29 (0.99–1.68)	**0.05**	1.31 (1.10–1.54)	**2.0E-03**	1.27 (1.08–1.51)	**4.0E-03**
CCPOC (high vs. low risk)	1.40 (1.08–1.81)	**0.01**	1.29 (1.00–1.69)	**0.04**	1.61 (1.39–1.88)	**3.4E-10**	1.55 (1.30–1.85)	**1.2E-06**

*One well-differentiated; two moderately differentiated; three poorly differentiated.
Numbers in bold indicate significance of 0.05 or less.

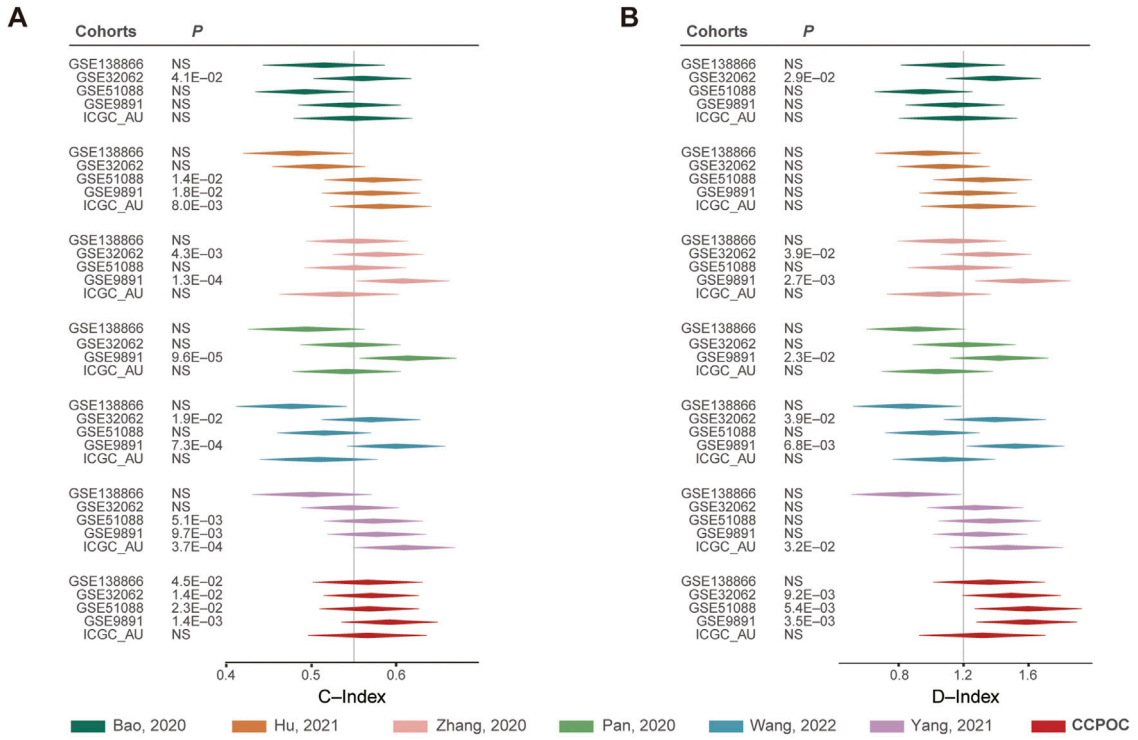
the risk score formula, patients were divided into CCPOC-high and CCPOC-low subgroups. Suboptimal samples were enriched in the CCPOC-high group; meanwhile, KIF20B was highly expressed in the CCPOC-high group, while POLA2 was highly expressed in the CCPOC-low group (Supplementary Figure S4). The CCPOC showed stronger prognostic efficiency than its individual constituents (Figure 3A). In the GSE26172 cohort, patients in the CCPOC-high group had significantly poorer OS than patients (Figure 3B, Supplementary Table S5). Moreover, the CCPOC-high group had significantly reduced OS compared to the CCPOC-low group in the seven public validation cohorts (Figure 3C–I, Supplementary Table S5) and our internal validation BL-OC cohort (Figure 3J, Supplementary Table S5). In addition, the CCPOC remains effective at discriminating survival after adjusting for clinical factors, including sex and de-bulking status ($p < 0.05$, Supplementary Figure S5). To test whether the CCPOC was an independent prognostic predictor, univariate and multivariate Cox regression analyses were conducted in the GSE26172 cohort and meta-validation of public cohorts. After adjusting for the clinicopathological parameters, the CCPOC remained independently prognostic (Table 2). Together, these findings indicated that the CCPOC was an independent prognostic signature.

Comparison with existing prognostic models

Next, in order to compare the prognostic value of the CCPOC with published OC prognostic models, referred to as Bao et al. (2020), Hu et al. (2021), Zhang et al. (2021), Pan's (Pan and Ma, 2020), Qi's (Ye et al., 2021), Wang et al. (2022), and Yang et al. (2021) prognostic system, based on the OS data of the GSE138866, GSE32062, GSE51088, GSE9891, and ICGC-AU cohorts, the C-index and D-index were calculated. As presented in Figure 4, the C-index was significantly higher in CCPOC than existing Bao and Wang models (Figures 4A, C). Like the C-index, the D-index was significantly higher in CCPOC than in most of the existing prognostic systems (Figures 4B, D).

Illustrating the immune microenvironment composition, dysregulated pathways, and drug sensitivity in CCPOC-low and CCPOC-high groups

Earlier, we showed that the CCPOC could help risk stratification of OC patients. Next, we explored the immune microenvironment composition in CCPOC-low and CCPOC-high groups. Immunomodulators have been classified into three types of molecules which include immune-inhibitors, immuno-stimulators, and major histocompatibility complex (MHC) molecules. The DNA damage response (DDR) refers to the process by which the cell maintains integrity of the genome after insult. The CCPOC-high group presented lower expression of MHC I/II molecular, immuno-inhibitor markers, immuno-stimulator markers (Figure 5A), and DDR markers (Figure 5B). The immune cell infiltration results showed that the CCPOC-high group was enriched with T cell CD4 memory resting cells (Figure 5C). Then, we conducted GSEA between the CCPOC-high and CCPOC-low groups (Supplementary Table S6). The EMT, TGF-β, and Wnt pathways were upregulated in the CCPOC-high group (Figure 6A). When analyzing CCPOC to predict chemotherapy sensitivity in the GSE146965 and PMID17290060 cohorts, chemotherapy effectiveness was lower in the CCPOC-high group than in the CCPOC-low group (Figure 6B). Immunotherapy, represented by a PD-L1/L1 blockade, has become a new breakthrough in cancer treatment. We analyzed the association between CCPOC and the response to immune checkpoint blockade therapy in two immunotherapy cohorts. In both the anti-PD-L1 cohort (IMvigor210) and the anti-PD-1 cohort (GSE78220), patients within the CCPOC-low group showed prolonged survival (Figures 6D,F). Treatment results showed that patients within the CCPOC-low group showed more efficacy against anti-PD-1/L1 immunotherapy than CCPOC-high group patients (Figure 6E and 6G-H). The aforementioned data indicate that the CCPOC-high group had lower tumor immunogenicity and lower efficacy of chemotherapy and anti-PD-1/L1 immunotherapy treatment response.



C

Meta C-index comparison

Signature	CCPOC vs. Other Signature	*Test of Superiority (P)*
CCPOC vs. Bao	0.57 *vs.* 0.53	1.00E-02
CCPOC vs. Hu	0.57 *vs.* 0.54	NS
CCPOC vs. Zhang	0.57 *vs.* 0.56	NS
CCPOC vs. Pan	0.57 *vs.* 0.55	NS
CCPOC vs. Wang	0.57 *vs.* 0.53	2.80E-02
CCPOC vs. Yang	0.57 *vs.* 0.56	NS

D

Meta D-index comparison

Signature	CCPOC vs. Other Signature	*Test of Superiority (P)*
CCPOC vs. Bao	1.48 *vs.* 1.15	5.00E-03
CCPOC vs. Hu	1.48 *vs.* 1.16	1.00E-02
CCPOC vs. Zhang	1.48 *vs.* 1.25	NS
CCPOC vs. Pan	1.48 *vs.* 1.13	1.10E-02
CCPOC vs. Wang	1.48 *vs.* 1.15	5.60E-03
CCPOC vs. Yang	1.48 *vs.* 1.24	3.80E-02

FIGURE 4
Comparison of published classifiers with CCPOC. Forest plot reporting of the **(A)** concordance index (C-index) and **(B)** D-index (robust hazard ratio) for CCPOC and published classifiers. The tables illustrate the superiority of CCPOC compared with published classifiers for Meta C-index **(C)** and Meta D-index **(D)**.

Exploring the prognostic significance and the immune therapy response in pan-cancer

Next, we explored the CCPOC signature in pan-cancer. As shown in Figure 7A, the CCPOC scores of OC were medium in the 33 cancers. Then, we explored the prognostic significance of CCPOC in pan-cancer. As presented in Figure 7B, CCPOC acted as a prognostic factor in almost 25% of cancer types. In addition, the CCPOC-high group presented low lymphocyte infiltration, such as CD8+ T cell, B cells, and NK cells (Figure 7C). Through an extensive literature search, we identified 15 tumor types for which data regarding the ORR of anti-PD-1/PD-L1 were available (OC excluded, Supplementary Table S7). In line with our suggestion, the CCPOC was negative to anti-PD-1/PD-L1 response in pan-cancer (r = −0.47, OC excluded) (Figure 7D). These data showed that CCPOC could be a biomarker in predicting the prognosis and anti-PD-1/PD-L1 response in other cancer types.

Discussion

Ovarian cancer (OC) is the most lethal gynecological cancer with pathological and molecular heterogeneity characteristics (Siegel et al., 2019). Various multi-gene prognostic biomarkers have been developed (Pan and Ma, 2020; Yang et al., 2021), but their prediction efficiencies are still uncertain. Therefore, a new signature that can accurately recognize OC patients with poor prognoses is urgently needed. Based on transcriptome data, OC has been unsupervised

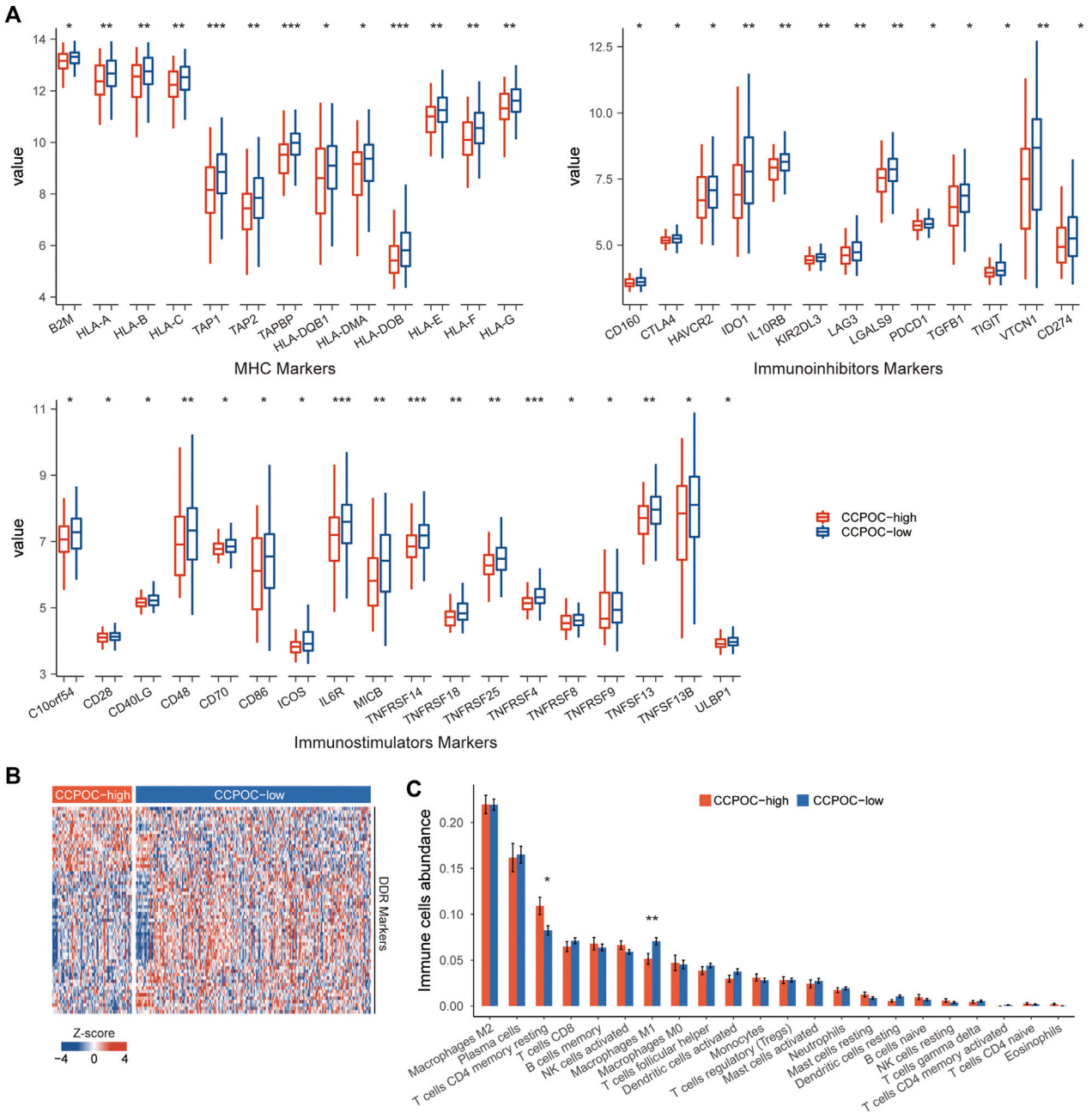


FIGURE 5
Immune microenvironment composition in CCPOC groups. **(A)** Expression levels of immuno-inhibitors, MHC I/II molecular, and immuno-stimulator markers within the CCPOC groups. **(B)** Heatmap of the expression of DDR markers between the CCPOC-high and CCPOC-low groups. **(C)** Distribution of 22 immune cells in the CCPOC groups. (*$p < 0.05$ and **$p < 0.01$).

and classified into four molecular subtypes (immunoreactive, differentiated, proliferative, and mesenchymal) with distinct molecular and clinical characteristics (Cancer Genome Atlas Research Network, 2011). The mesenchymal subtype highly expressed the mesenchymal signature and was associated with worse clinical outcomes. The prognostic signature screened based on molecular portraits specific to the worst prognosis subtype may be used for risk stratification of OC patients. In addition, many studies have explored the role of cell cycle in the prognosis prediction of tumors (Hui et al., 2021; Jiang et al., 2021). However, most studies have only studied the prognostic relevance of cell cycle without considering tumor heterogeneity.

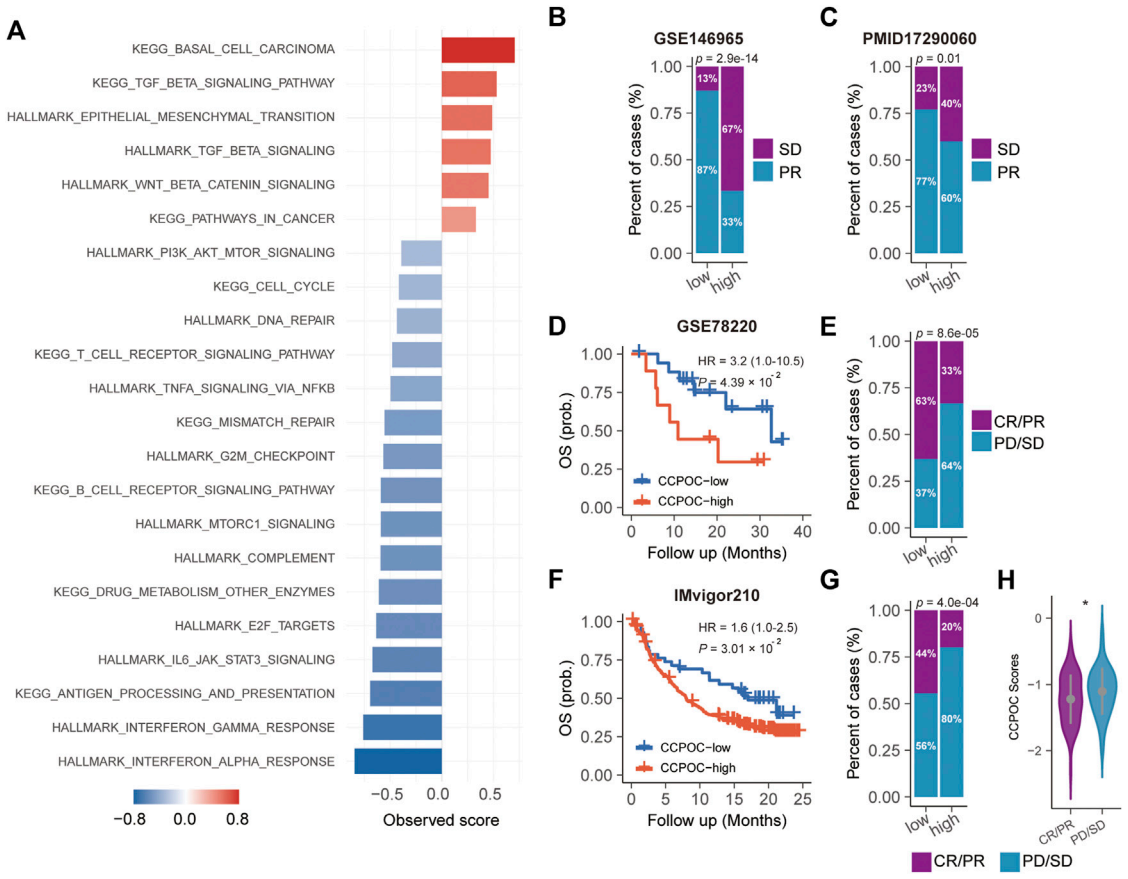


FIGURE 6
Dysregulated pathways and drug sensitivity within the CCPOC groups. **(A)** Enrichment of dysregulated pathways between the high- and low-CCPOC groups. **(B)** and **(C)** chemotherapy sensitivity analysis. Survival analyses for the CCPOC groups in the GSE78220 cohort **(D)** and the IMvigor210 cohort **(F)**. Proportion of patients with response to anti-PD-1/L1 immunotherapy in low- or high-CCPOC groups in the GSE78220 cohort **(E)** and the IMvigor210 cohort **(G)**. **(H)** Distribution of the CCPOC scores in CCPOC groups. CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease.

In this study, we applied a network-based approach to integrate cell cycle signature and modalities underlying the mesenchymal subtype to establish a prognostic signature termed "Cell Cycle Prognostic Signature of Ovarian Cancer" (CCPOC). To our knowledge, no prognostic cell cycle-based signature has been constructed by incorporating molecular subtyping information of OC.

The CCPOC was constructed by two cell cycle genes (POLA2 and KIF20B) which were key regulators of the mesenchymal subtype and could stratify patients into different risk groups. Within these two cell cycle genes, KIF20B can promote cell proliferation and could be a potential therapeutic target in pancreatic cancer (Chen et al., 2021). Koh et al. reported that the knockdown of POLA2 increases chemo-resistance in human lung cancer cells (Koh et al., 2016). The defined CCPOC-high group showed a worse OS than the CCPOC-low group. To confirm this finding, we validated the results in seven independent cohorts measured by various platforms and one independent internal cohort (BL-OC cohort) and found that the signature successfully stratified the prognosis in all cohorts. The CCPOC remained an independent prognostic predictor in the multivariate Cox proportional hazards analysis after adjusting for other clinical factors. In line with the findings, we found that the C-index and D-index of the CCPOC were significantly higher than those of the published prognostic models, which was superior to the current genomic classification. These data suggest that the CCPOC has a strong and reproducible prognostic value for risk stratification of OC. In addition, we also found that the CCPOC was related to weakened tumor immunogenicity and inflamed antitumor immunity, and the correlation analysis showed that CCPOC was negatively related to the ORR in pan-cancer (OC excluded), which indicated that the CCPOC-low group may be sensitive to anti-PD-1/PD-L1 therapy. Together, these findings show that the CCPOC could serve as a robust prognostic signature in OC.

This study still has some limitations. First, the prognostic signature was screened from gene expression profiles generated from microarray platforms, which are expensive, difficult to

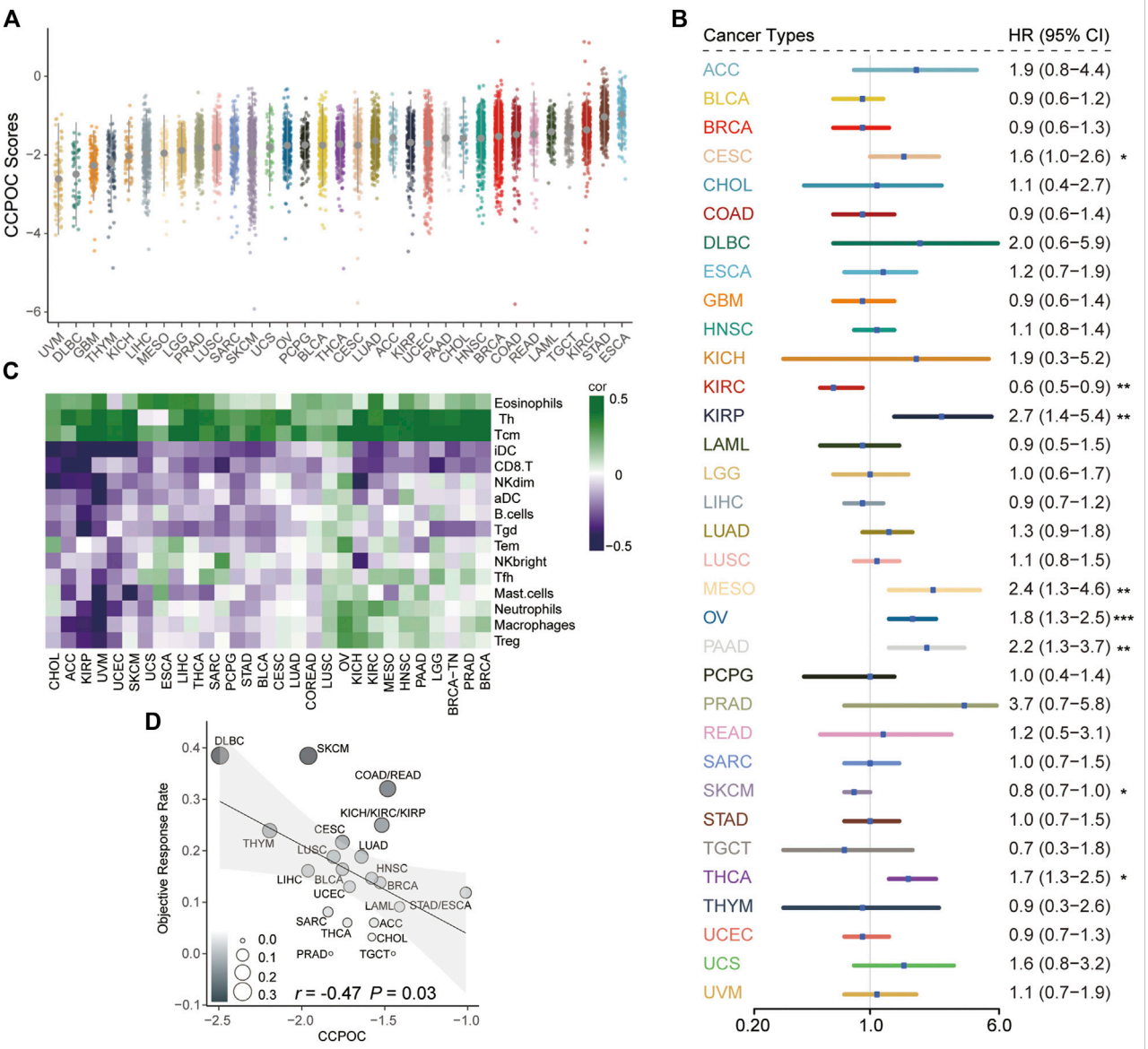


FIGURE 7
Pan-cancer analysis of the prognostic significance and anti-PD-1/PD-L1 response for CCPOC. **(A)** Distribution of CCPOC scores in pan-cancer. **(B)** Prognostic relevance of CCPOC in various cancer types. **(C)** Heatmap of the correlation of immune cells and CCPOC scores across various cancer types. **(D)** Correlation between CCPOC and anti-PD-1/PD-L1 ORR of various cancer types. (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$).

operate, and involve professional bioinformatics expertise, so it is difficult to be popularized in daily clinical application. Second, the training and validation datasets were all from retrospective studies in the study, including fresh frozen samples. Therefore, in practice, we need to detect the expression of signature genes using conventional clinical techniques, such as RT-PRC or IHC, and then reconstruct the new model and perform large-scale multicenter cohorts to validate the validity and robustness of the model.

In conclusion, using multi-dimensional network inference underlying the mesenchymal subtype of OC, we have identified and validated a two cell cycle signature, named CCPOC, to risk-stratify patients and provide an easy method for the exploration of new effective therapeutic options, including novel target drugs and immune therapy in the future.

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.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the Beilun People's Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LC and YH contributed to the conception and design of the project. LC, HG, JW, and LZ participated in the data collection and data analyses. YH and CS wrote and edited the manuscript. All the authors read and approved the final manuscript. All authors contributed to the manuscript and approved the submitted version.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.958092/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Kaplan–Meier analysis of different molecular subtypes.

SUPPLEMENTARY FIGURE S2
Expression levels of POLA2 and KIF20B **(B)** genes among different subtypes.

SUPPLEMENTARY FIGURE S3
2-year and 5-year AUC of the CCPOC score in the BL-OC cohort.

SUPPLEMENTARY FIGURE S4
Heatmap of the expression of the two candidate immune genes.

SUPPLEMENTARY FIGURE S5
Kaplan–Meier analysis of CCPOC risk groups within different conditions.

References

Bao, M., Zhang, L., and Hu, Y. (2020). Novel gene signatures for prognosis prediction in ovarian cancer. *J. Cell. Mol. Med.* 24 (17), 9972–9984. doi:10.1111/jcmm.15601

Bertoli, C., Skotheim, J. M., and De Bruin, R. A. (2013). Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* 14 (8), 518–528. doi:10.1038/nrm3629

Bonome, T., Levine, D. A., Shih, J., Randonovich, M., Pise-Masison, C. A., Bogomolniy, F., et al. (2008). A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res.* 68 (13), 5478–5486. doi:10.1158/0008-5472.CAN-07-6595

Bottoni, P., and Scatena, R. (2015). The role of CA 125 as tumor marker: biochemical and clinical aspects. *Adv. Exp. Med. Biol.* 867, 229–244. doi:10.1007/978-94-017-7215-0_14

Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474 (7353), 609–615. doi:10.1038/nature10166

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Chen, G. M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., Gendoo, D. M., et al. (2018). Consensus on molecular subtypes of high-grade serous ovarian carcinoma. *Clin. Cancer Res.* 24 (20), 5037–5047. doi:10.1158/1078-0432.CCR-18-0784

Chen, J., Zhao, C.-C., Chen, F.-R., Feng, G.-W., Luo, F., and Jiang, T. (2021). KIF20B promotes cell proliferation and may Be a potential therapeutic target in pancreatic cancer. *J. Oncol.* 2021, 5572402. doi:10.1155/2021/5572402

Coleman, R. L., Monk, B. J., Sood, A. K., and Herzog, T. J. (2013). Latest research and treatment of advanced-stage epithelial ovarian cancer. *Nat. Rev. Clin. Oncol.* 10 (4), 211–224. doi:10.1038/nrclinonc.2013.5

Cuzick, J., Swanson, G. P., Fisher, G., Brothman, A. R., Berney, D. M., Reid, J. E., et al. (2011). Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet. Oncol.* 12 (3), 245–255. doi:10.1016/S1470-2045(10)70295-3

Denkert, C., Budczies, J., Darb-Esfahani, S., Györffy, B., Sehouli, J., Könsgen, D., et al. (2009). A prognostic gene expression index in ovarian cancer—validation across different independent data sets. *J. Pathol.* 218 (2), 273–280. doi:10.1002/path.2547

Dressman, H. K., Berchuck, A., Chan, G., Zhai, J., Bild, A., Sayer, R., et al. (2007). An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J. Clin. Oncol.* 25 (5), 517–525. doi:10.1200/JCO.2006.06.3743

Fletcher, M. N., Castro, M. A., Wang, X., De Santiago, I., O'Reilly, M., Chin, S.-F., et al. (2013). Master regulators of FGFR2 signalling and breast cancer risk. *Nat. Commun.* 4 (1), 2464–2512. doi:10.1038/ncomms3464

Hu, Y., Taylor-Harding, B., Raz, Y., Haro, M., Recouvreux, M. S., Taylan, E., et al. (2020). Are Epithelial ovarian cancers of the mesenchymal subtype actually intraperitoneal metastases to the ovary? *Front. Cell Dev. Biol.* 8, 647. doi:10.3389/fcell.2020.00647

Hu, Y., Zheng, M., Wang, S., Gao, L., Gou, R., Liu, O., et al. (2021). Identification of a five-gene signature of the RGS gene family with prognostic value in ovarian cancer. *Genomics* 113 (4), 2134–2144. doi:10.1016/j.ygeno.2021.04.012

Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hu-Lieskovan, S., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 165 (1), 35–44. doi:10.1016/j.cell.2016.02.065

Hui, Y., Leng, J., Jin, D., Liu, D., Wang, G., Wang, Q., et al. (2021). A cell cycle progression-derived gene signature to predict prognosis and therapeutic response in hepatocellular carcinoma. *Dis. Markers* 2021, 1986159. doi:10.1155/2021/1986159

Jiang, W., Xu, J., Liao, Z., Li, G., Zhang, C., and Feng, Y. (2021). Prognostic signature for lung adenocarcinoma patients based on cell-cycle-related genes. *Front. Cell Dev. Biol.* 9, 635. doi:10.3389/fcell.2021.655950

Jiménez-Sánchez, A., Cybulska, P., Mager, K. L., Koplev, S., Cast, O., Couturier, D.-L., et al. (2020). Unraveling tumor–immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy. *Nat. Genet.* 52 (6), 582–593. doi:10.1038/s41588-020-0630-5

Karlan, B. Y., Dering, J., Walsh, C., Orsulic, S., Lester, J., Anderson, L. A., et al. (2014). POSTN/TGFBI-associated stromal signature predicts poor prognosis in serous epithelial ovarian cancer. *Gynecol. Oncol.* 132 (2), 334–342. doi:10.1016/j.ygyno.2013.12.021

Knijnenburg, T. A., Wang, L., Zimmermann, M. T., Chambwe, N., Gao, G. F., Cherniack, A. D., et al. (2018). Genomic and molecular landscape of DNA damage repair deficiency across the Cancer Genome Atlas. *Cell Rep.* 23 (1), 239e6–254. doi:10.1016/j.celrep.2018.03.076

Koh, V., Kwan, H. Y., Tan, W. L., Mah, T. L., and Yong, W. P. (2016). Knockdown of POLA2 increases gemcitabine resistance in lung cancer cells. *BMC genomics* 17 (13), 1029–1096. doi:10.1186/s12864-016-3322-x

Konecny, G. E., Wang, C., Hamidi, H., Winterhoff, B., Kalli, K. R., Dering, J., et al. (2014). Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J. Natl. Cancer Inst.* 106 (10), dju249. doi:10.1093/jnci/dju249

Lheureux, S., Gourley, C., Vergote, I., and Oza, A. M. (2019). Epithelial ovarian cancer. *Lancet* 393 (10177), 1240–1253. doi:10.1016/S0140-6736(18)32552-2

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

Mo, X., Su, Z., Yang, B., Zeng, Z., Lei, S., and Qiao, H. (2020). Identification of key genes involved in the development and progression of early-onset colorectal cancer by co-expression network analysis. *Oncol. Lett.* 19 (1), 177–186. doi:10.3892/ol.2019.11073

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Pan, X., and Ma, X. (2020). A novel six-gene signature for prognosis prediction in ovarian cancer. *Front. Genet.* 11, 1006. doi:10.3389/fgene.2020.01006

Patch, A.-M., Christie, E. L., Etemadmoghadam, D., Garsed, D. W., George, J., Fereday, S., et al. (2015). Whole–genome characterization of chemoresistant ovarian cancer. *Nature* 521 (7553), 489–494. doi:10.1038/nature14410

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA. A Cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551

Tamborero, D., Rubio-Perez, C., Muiños, F., Sabarinathan, R., Piulats, J. M., Muntasell, A., et al. (2018). A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clin. Cancer Res.* 24 (15), 3717–3728. doi:10.1158/1078-0432.CCR-17-3509

Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., et al. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* 14 (16), 5198–5208. doi:10.1158/1078-0432.CCR-08-0196

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17 (1), 98–110. doi:10.1016/j.ccr.2009.12.020

Wang, X., Terfve, C., Rose, J. C., and Markowetz, F. (2011). HTSanalyzeR: An R/bioconductor package for integrated network analysis of high-throughput screens. *Bioinformatics* 27 (6), 879–880. doi:10.1093/bioinformatics/btr028

Wang, Y., Li, B. X., and Li, X. (2022). Identification and validation of angiogenesis-related gene expression for predicting prognosis in patients with ovarian cancer. *Front. Oncol.* 11, 783666. doi:10.3389/fonc.2021.783666

Williams, G. H., and Stoeber, K. (2012). The cell cycle and cancer. *J. Pathol.* 226 (2), 352–364. doi:10.1002/path.3022

Yang, J., Hong, S., Zhang, X., Liu, J., Wang, Y., Wang, Z., et al. (2021). Tumor immune microenvironment related gene-based model to predict prognosis and response to compounds in ovarian cancer. *Front. Oncol.* 11, 807410. doi:10.3389/fonc.2021.807410

Ye, Y., Dai, Q., Li, S., He, J., and Qi, H. (2021). A novel defined risk signature of the ferroptosis-related genes for predicting the prognosis of ovarian cancer. *Front. Mol. Biosci.* 8, 645845. doi:10.3389/fmolb.2021.645845

Yoshihara, K., Tsunoda, T., Shigemizu, D., Fujiwara, H., Hatae, M., Fujiwara, H., et al. (2012). High-risk ovarian cancer based on 126-gene expression signature is uniquely characterized by downregulation of antigen presentation pathway. *Clin. Cancer Res.* 18 (5), 1374–1385. doi:10.1158/1078-0432.CCR-11-2725

Zhang, D., Li, Y., Yang, S., Wang, M., Yao, J., Zheng, Y., et al. (2021). Identification of a glycolysis-related gene signature for survival prediction of ovarian cancer patients. *Cancer Med.* 10 (22), 8222–8237. doi:10.1002/cam4.4317

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 19 August 2022
DOI 10.3389/fmolb.2022.951432

Check for updates

OPEN ACCESS

EDITED BY
Hem Chandra Jha,
Indian Institute of Technology Indore, India

REVIEWED BY
Pedro José Carlos Rondot Radío, University of Buenos Aires, Argentina
Sampath Chandra Prasad Rao, Bangalore Skull Base Insititue, India

*CORRESPONDENCE
Juehui Wang,
m13886588857@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 24 May 2022
ACCEPTED 27 July 2022
PUBLISHED 19 August 2022

CITATION
Mao C, Li A, Hu J, Wang P, Peng D, Wang J and Sun Y (2022), Efficient and accurate diagnosis of otomycosis using an ensemble deep-learning model. *Front. Mol. Biosci.* 9:951432. doi: 10.3389/fmolb.2022.951432

COPYRIGHT
© 2022 Mao, Li, Hu, Wang, Peng, Wang 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.

Efficient and accurate diagnosis of otomycosis using an ensemble deep-learning model

Chenggang Mao[1†], Aimin Li[2†], Jing Hu[3], Pengjun Wang[4], Dan Peng[1], Juehui Wang[5*] and Yi Sun[3]

[1]Department of Otolaryngology Head and Neck Surgery, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, Hubei, China, [2]Department of Pediatrics, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, Hubei, China, [3]Department of Dermatology, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, Hubei, China, [4]Health Science Center, Yangtze University, Jingzhou, Hubei, China, [5]Department of Information Engineering, Jingzhou University, Jingzhou, Hubei, China

Otomycosis accounts for over 15% of cases of external otitis worldwide. It is common in humid regions and Chinese cultures with ear-cleaning custom. *Aspergillus* and *Candida* are the major pathogens causing long-term infection. Early endoscopic and microbiological examinations, performed by otologists and microbiologists, respectively, are important for the appropriate medical treatment of otomycosis. The deep-learning model is a novel automatic diagnostic program that provides quick and accurate diagnoses using a large database of images acquired in clinical settings. The aim of the present study was to introduce a machine-learning model to accurately and quickly diagnose otomycosis caused by *Aspergillus* and *Candida*. We propose a computer-aided decision-making system based on a deep-learning model comprising two subsystems: Java web application and image classification. The web application subsystem provides a user-friendly webpage to collect consulted images and display the calculation results. The image classification subsystem mainly trained neural network models for end-to-end data inference. The end user uploads a few images obtained with the ear endoscope, and the system returns the classification results to the user in the form of category probability values. To accurately diagnose otomycosis, we used otoendoscopic images and fungal culture secretion. Fungal fluorescence, culture, and DNA sequencing were performed to confirm the pathogens *Aspergillus* or *Candida* spp. In addition, impacted cerumen, external otitis, and normal external auditory canal endoscopic images were retained for reference. We merged these four types of images into an otoendoscopic image gallery. To achieve better accuracy and generalization abilities after model-training, we selected 2,182 of approximately 4,000 ear endoscopic images as training samples and 475 as validation samples. After selecting the deep neural network models, we tested the ResNet, SENet, and EfficientNet neural network models with different numbers of layers. Considering the accuracy and operation speed, we finally chose the EfficientNetB6 model, and the probability values of the four categories of otomycosis, impacted cerumen, external otitis, and normal cases were outputted. After multiple model training iterations, the average accuracy of the overall validation sample reached 92.42%. The results suggest that the system could be used as a reference for general practitioners to obtain more accurate diagnoses of otomycosis.

KEYWORDS

otomycosis, *Aspergillus*, *Candida*, deep-learning, otoendoscopic

1 Introduction

The goal of machine learning is for machines to learn automatically from training data and update their capabilities (Shen et al., 2017). Deep learning is a field of machine learning in which machine learning is implemented using deep neural networks (Giger, 2018). For example, the deep convolutional neural network (CNN) is used by machines to acquire image data and identify the contents in the image (Kooi et al., 2017). Deep learning, or deep neural networks, has been successfully used in some medical fields. For example, in the field of ophthalmology, machine learning can automatically detect retinopathy in patients with diabetes (Horsch et al., 2011; Chougrad et al., 2018).

Otomycosis externa, or fungal otitis externa, is a superficial fungal infection of the external auditory canal that occasionally invades the middle ear. The incidence of otomycosis externa is high in hot and humid climates found in tropical and subtropical regions (Li and He 2019). Patients with this infection usually present with symptoms such as itching, otorrhea, and hearing loss. *Aspergillus* and *Candida* are the most common pathogens causing otomycosis (Kamali Sarwestani et al., 2018). Although otomycosis is rarely fatal, it is difficult to treat because of a long treatment period and easy recurrence. Antifungal drugs show varying sensitivity/resistance to pathogenic bacteria, often leading to poor therapeutic effects (Li et al., 2020). The diagnosis of otomycosis is mainly done clinically by otorhinolaryngologists, and insufficient attention is paid to mycological detection, particularly the subsequent pathogen culture and drug sensitivity test results.

To date, the deep CNN model with various structures has achieved good results in image classification and recognition (Hapfelmeier and Horsch, 2011; Lee et al., 2017). However, the tests were based on standard image galleries and have not been applied in the field of otoendoscopy.

The aim of this study was to establish a comprehensive identification system for otomycosis by comparing the effects of several typical deep CNN models using otoendoscopic images and web applications and selecting three models with the best recognition effects. To diagnose otomycosis accurately, we used otoendoscopic images and fungal culture secretions. The pathogen was identified as *Aspergillus* or *Candida* by fungal fluorescence, culture, and DNA sequencing. In addition, images of impacted cerumen, external otitis, and normal external auditory canal were used as reference to identify otomycosis. We combined these four types of images into a gallery of otoendoscopic images.

2 Materials and methods

2.1 Design

Figure 1 summarizes the overall design of the web-based computer-aided diagnosis system used in this study. The system comprises three subsystems: front-end page subsystem based on React, business logic subsystem based on SpringBoot, and image classification subsystem based on PyTorch. React is a JavaScript language library for building user interfaces, enabling component-based user interaction pages for easy extension. SpringBoot is a Java language-based framework for quickly building standalone, production-level Java Web services applications. PyTorch is an open-source framework for deep learning based on Python and developed with support from Facebook (Zhang et al., 2019).

According to manufacturer's instructions, we uploaded a batch of otoendoscopic RGB images up to four at a time to the specific storage space of the server *via* the front-end subsystem. Subsequently, the business logic subsystem verified and cleaned the images. Finally, the image classification subsystem predicted the images for otomycosis, impacted cerumen, external otitis, and the normal external auditory canal. At the end of the model calculation, each image was assigned a percentage of the four categories, adding up to 100%, with the value of each category indicating the probability that the image belonged to that category. Finally, the data were delivered to the front-end page *via* the business subsystem using web services to provide users with diagnostic references. To classify and predict RGB images, we used an end-to-end deep CNN model (Esteva et al., 2017). To train weight parameters, we inputted the classified and labeled training sample images to the model.

2.2 Sample

The database used in our study was created at the Department of Otolaryngology - Head and Neck Surgery, Jingzhou Hospital Affiliated to Yangtze University. The study protocol was approved by the Research Ethics Committee of Jingzhou Hospital Affiliated to Yangtze University (protocol number: 2021-093-01). Written informed consent was obtained from patients and caregivers of patients under 18 years of age. All the procedures were carried out in accordance with the tenets of the Declaration of Helsinki.

An otorhinolaryngologist evaluated the patients. Patient age ranged from 5 to 72 years. Images with otomycosis, impacted cerumen, external otitis, or the normal external auditory canal

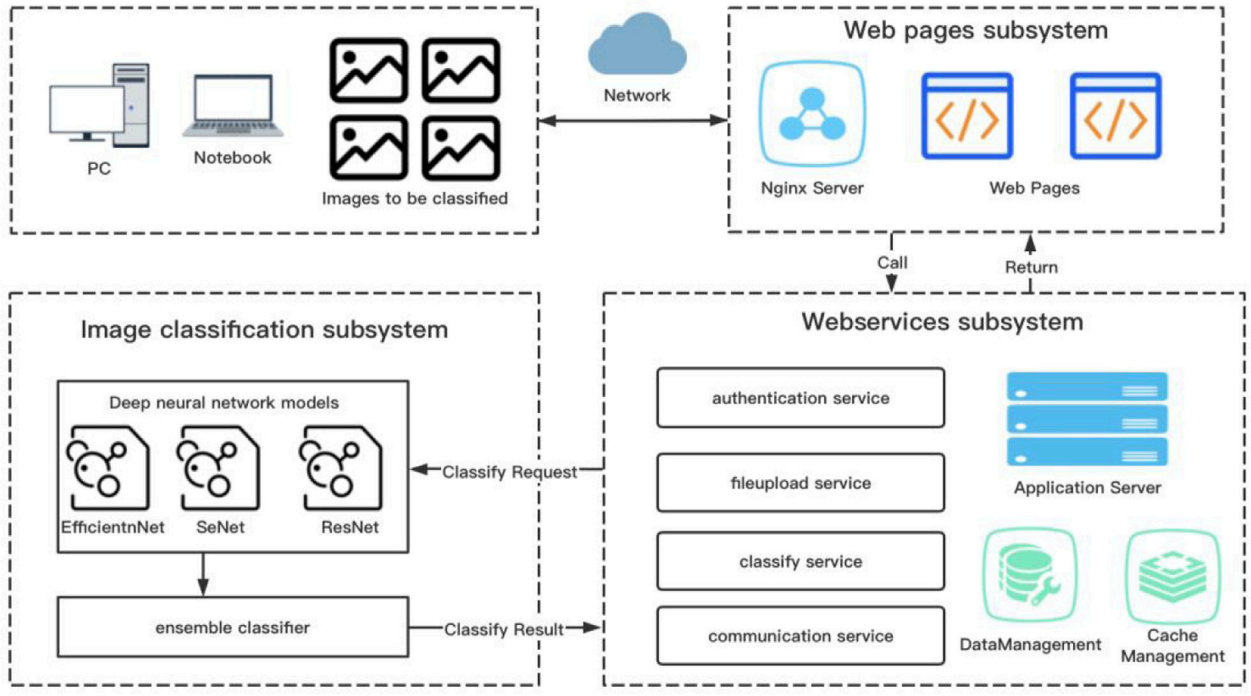


FIGURE 1
General scheme of the computer-aided system approach to assist the diagnosis of otomycosis.

were preserved in the otoscopy room. Figures 2A–D shows the representative images of all categories.

To obtain a better training result, we prepared two external auditory canal image datasets: one comprising 2,652 external auditory canal images used to train the model and the other comprising 552 images that were not used to train the model but to test the robustness and accuracy of the model.

2.3 Data preprocessing

First, we selected the otoendoscopic RGB image as the system input. The key contents to be identified were in a circular area in the image, while the other areas were black. Due to inter-user differences in operation habits, the circular area varied in size, with the center of the circular area often not coinciding with the center of the image. To maximize the feature data to be recognized by the model for training, we used Hough transform to perform ring detection (Hough ring detection). The center and radius of the circle were detected to calculate the minimum rectangular coordinates. From the rectangular cutting of the original image, excess black boxes were removed (Moses et al., 2018).

Second, because the model was in the training process, the dataset was reused in each iteration of the training. To enhance the robustness of the model, we added the function of random angle rotation to the dataset (Varma and Zisserman, 2009) such that when each sample was removed, the image was rotated randomly.

Third, data standardization was performed. To accelerate the training process, the possibility of the model falling into the local optimum during training was reduced. We treated the input data for data standardization, i.e., according to the image color channel for the unit, the mean and standard deviation values of the training images were calculated. After normalization of each image pixel value, the mean value was subtracted and divided by the standard deviation value. The images with a mean value of 0 and a standard deviation of one were considered normally distributed (Shin et al., 2016). Due to limited memory, 600 training sample otoendoscopic images were randomly selected. The calculated mean value of the RGB channel was [0.5317, 0.3899, 0.3003], while the standard deviation was [0.3482, 0.2748, 0.2329]. The formula for the standard deviation was as follows: $X' = \frac{x-\mu}{\sigma}$

2.4 Deep CNN model

The deep CNN model was used in this study. It could analyze the features in its "field of vision" (local receptive field) through

10.3389/fmolb.2022.951432

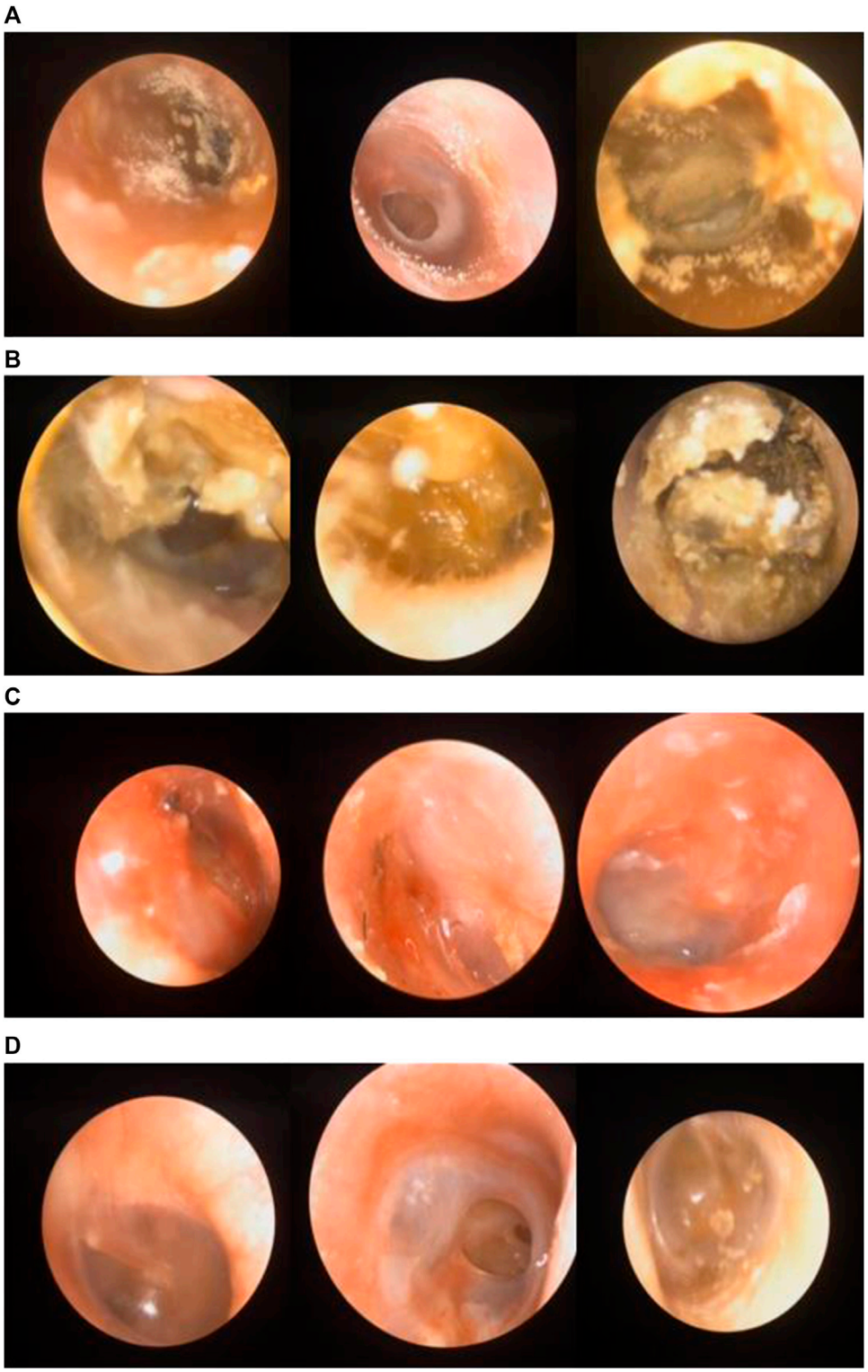


FIGURE 2
(A) Images labeled "otomycosis". **(B)** Images labeled "impacted cerumen". **(C)** Images labeled "external otitis". **(D)** Images labeled "normal case".

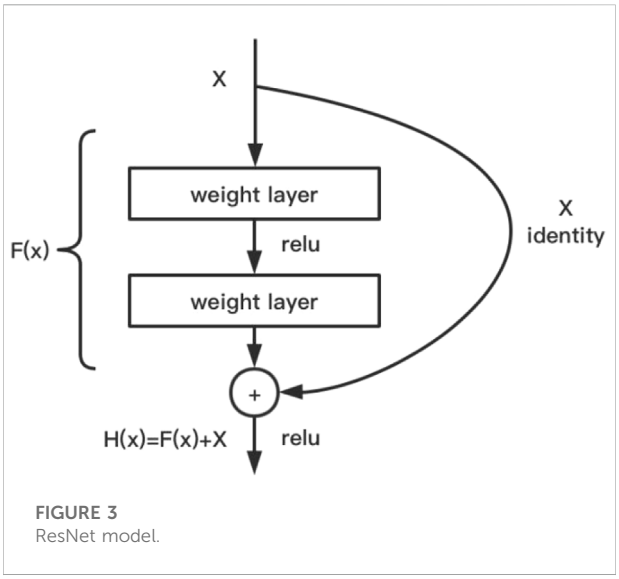


FIGURE 3
ResNet model.

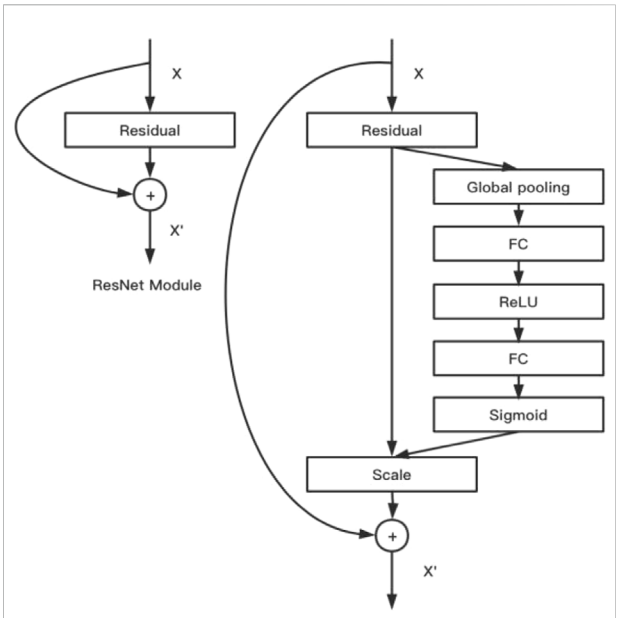


FIGURE 4
SENet model.

neuronal learning of the hidden layer. To enable the CNN model to learn better features from images, many deep CNN models have been proposed (Spasov et al., 2018). We selected three CNN models for training: ResNet, SENet, and EfficientNet (Lee et al., 2020). The algorithm characteristics and design ideas of each model are described below:

2.4.1 ResNet

The main design idea is to introduce a residual network structure (cross-layer jump connection) to resolve SGD optimization difficulties when the neural network model is stacked to a deeper level, and the reverse derivative gradient disappears or explodes, resulting in the deterioration of model performance (Talo et al., 2019). Using the neural network model with a residual structure, the network can be designed deeper, and the training is faster. Because it does not introduce additional parameters or computational complexities and only performs simple addition operation, the computational power consumption is negligible compared to convolution operation. ResNet has designed floors of 18, 34, 50, 101, and 152. Regarding efficiency and cost, ResNet of 101 was used in this study. Figure 3 shows the algorithm and design of this model.

2.4.2 SENet

This conventional CNN model aggregates information of image space and feature dimension into a feature channel through the convolution operation of multiple convolution kernels in the local receptive field, and the data of each feature channel are equal (Zhang et al., 2019). However, in real settings, each image is a key feature area, and attention should be paid to the channel data of this part of data conversion. The SE module was designed by the makers of SENet to obtain scaling coefficients (importance) of the channel data through a series of fully connected activation operations of global average pooling and 1 × 1 convolution kernels, which were then weighted to previous features by multiplication. This completed the recalibration of the original feature on the channel dimension. The SE module and ResNet were combined to obtain SENet. In this study, we used the 101-layer SENet. Figure 4 shows the algorithm and design of this model.

2.4.3 EfficientNet

Scaling (model extension) improves the performance of CNN models (Godinez et al., 2017). The expansion direction of the model is mainly divided into the network width (channel, number of convolution kernels), depth (number of layers), and resolution (accuracy of input image); however, these expansions consume abundant additional computing resources. EfficientNetB0 is a simple baseline network with grid structure search. Subsequently, a compound coefficient is used to synthesize the aforementioned dimensions of model extension such that additional consumption of computational power resources can be optimized to improve accuracy. In this study, we used EfficientNetB6.

2.5 Deep CNN model training process

2.5.1 Input and output

The size of otoendoscopic RGB images was 515 × 547 pixels, and the standard input size of ResNet101 and SENet101 models was 224 × 224 pixels. Therefore, we added a convolution layer in

TABLE 1 Highest accuracies of the four categories of otoendoscopic images trained by three models.

	Otomycosis (%)	Impacted cerumen (%)	External otitis (%)	Normal case (%)
ResNet101	73.8	78.69	71.19	86.75
SENet101	89.25	78.69	83.33	89.82
EfficientNetB6	95.19	78.69	80.0	86.83

TABLE 2 Weighted mean values of the models for four categories of otoendoscopic images.

Weighted mean model	Otomycosis	Impacted cerumen	External otitis	Normal case
ResNet101	0.28578066	0.33333333	0.3035562	0.329347
SENet101	0.34560873	0.33333333	0.35532151	0.34100227
EfficientNetB6	0.3686106	0.33333333	0.3411223	0.32965073

front of the input layer of the standard ResNet101 and SENet101. The kernel size was 7 × 7. The stride and padding were two and three, respectively, such that the input size of the model became 448 × 448. The input size of the EfficientNetB6 model was 528 × 528. Therefore, the image was scaled to the required size of the model before training. To adapt the output of the model to the objective of our study, the otoendoscopic images were classified into four categories. We replaced the last full connection layer of the standard ResNet101, SENet101, and EfficientNetB6 models with a new full connection layer with four output nodes (Geras et al., 2019).

2.5.2 Optimization of model training

To obtain better training results, we assigned a set of hyperparameters, including whether or not to use pretraining weights, whether or not to carry out center circle interception, batch size, learning rate, and optimizer (Zhang and Suganthan, 2017).

Each time the model was trained, a value was randomly selected from each item. We used the PyTorch framework on two graphics processing units (RTX-2080Ti Gpus). After repeated training of multiple models, we obtained the optimal hyperparameters: using pretraining weight, using center circle interception, batch size of 4, learning rate of 0.001, and SGD optimizer.

2.6 Set classifier

Based on the accuracy of the test images, the best training results from among the three models were selected. The average highest accuracies of the ResNet101, SENet101, and EfficientNetB6 models were 78.32, 87.16, and 88.21%, respectively. Table 1 shows the highest accuracies for the four categories. Table 2 shows the weighted mean values of the models for the four categories. The set classifier was obtained by multiplying each model by the weighted mean of each model and category. Table 3 shows the highest accuracy of the final set classifier.

The performance of the set classifier was measured using the obfuscation matrix, precision, recall, precision–recall (PR) curve, and receiver operating characteristic (ROC) curve (Tan et al., 2017).

The confusion matrix represented the number of instances corresponding to the predicted and actual classes. This concept is often used for binary classifications but can be extended to multiclass predictions, with the corresponding class on the diagonal of the matrix and the misclassified class outside the diagonal. We used the set classifier to conduct prediction tests on the verification test set comprising 475 otoendoscopic images. Table 4 shows the confusion matrix.

The accuracy refers to the number of predicted positive sample results that are correctly classified. It is calculated using the following formula: precision = true positive/(true positive + false positive). Recall refers to the number of positive sample results that are correctly classified. It is calculated using the following formula: recall = true positive/(true positive + false negative). Table 5 shows the results of the precision and recall.

The PR curve is used to sort the samples according to the predicted results of the classifier. The samples considered “most likely” to be positive by the classifier were in the front row, while those considered “least likely” to be positive by the classifier were in the back row. In this order, the samples were considered examples of positive prediction, and the current recall and precision were calculated each time. With accuracy as the vertical axis and

TABLE 3 Highest accuracies of four categories of otoendoscopic images by the ensemble classifier.

	Otomycosis (%)	Impacted cerumen (%)	External otitis (%)	Normal case (%)	Average accuracy (%)
Set classifier	94.65	90.16	88.33	92.22	92.42

Evaluation index.

TABLE 4 Confounding matrix results of the test set verified on otoendoscopic images.

Prediction fact	Otomycosis	Impacted cerumen	External otitis	Normal case
Otomycosis	177	3	4	3
Impacted cerumen	2	55	0	4
External otitis	1	0	53	6
Normal case	5	1	7	154

TABLE 5 Accuracy and recall of the four categories of otoendoscopic images.

Indicators	Category			
	Otomycosis (%)	Impacted cerumen (%)	External otitis (%)	Normal case (%)
Precision	95.68	93.22	82.81	92.22
Recall	94.65	90.16	88.33	92.22

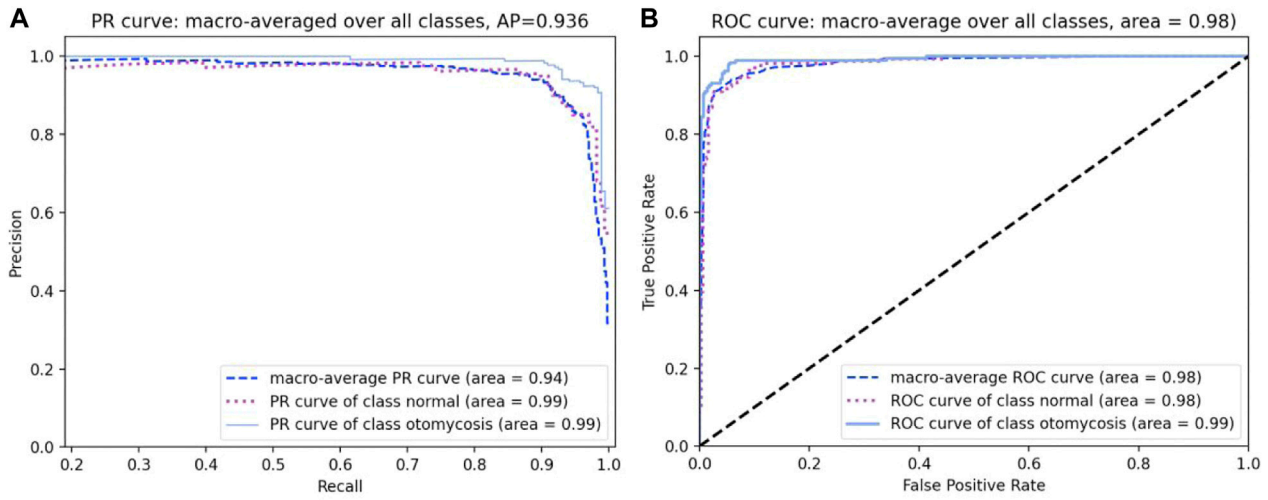


FIGURE 5
(A,B) Precision–recall and receiver operating characteristic curves of the four categories of otoendoscopic images.

recall as the horizontal axis, the PR curve was drawn. With the true-positive rate as the vertical axis and the false-positive rate as the horizontal axis, the ROC curve was drawn. Figures 5A,B show the results of the PR and ROC curves, respectively.

3 Results

We chose a deep CNN model to obtain a system that could help doctors accurately diagnose otomycosis. Three CNN models

TABLE 6 Data distribution in the classification test of otoendoscopic images.

Classification	Training set	Validation set	Test set	Total
Otomycosis	803	187	30	1,020
Impacted cerumen	264	61	30	355
External otitis	395	60	30	485
Normal case	720	167	30	917

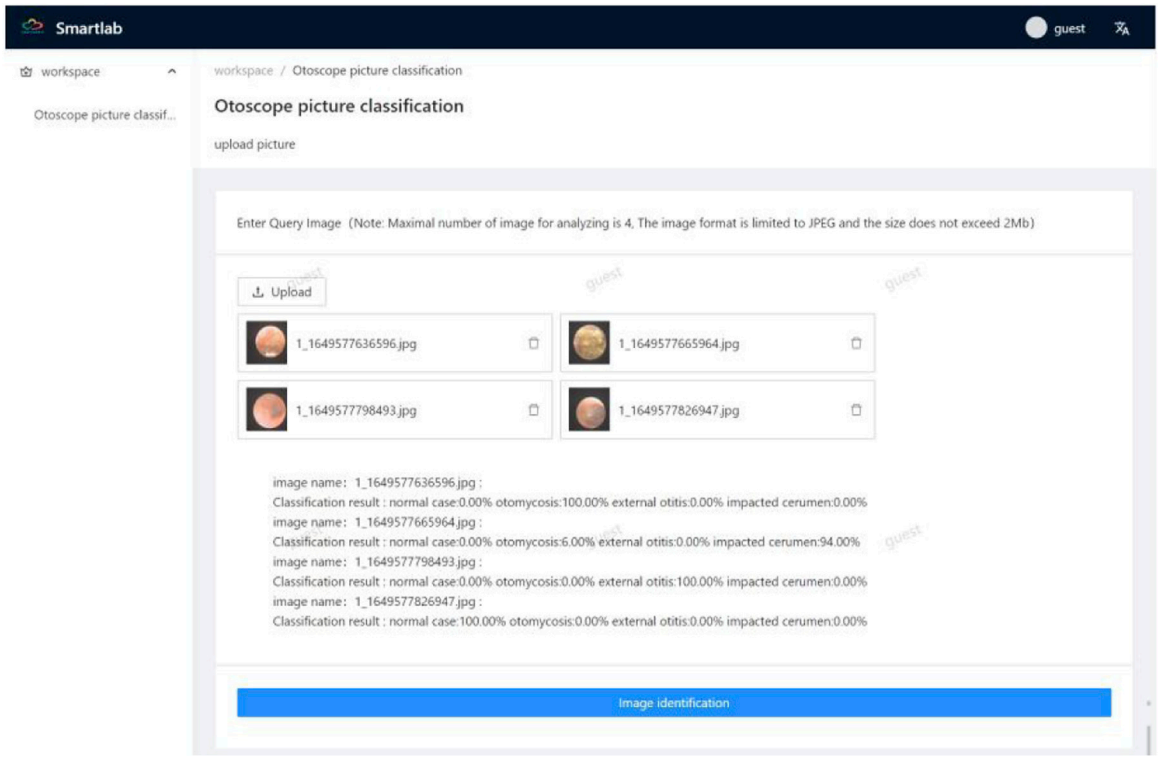


FIGURE 6
Screenshot of the authentication results on the webpage.

were selected for training, and the weighted mean of each model produced the highest verification accuracy.

To obtain better accuracy and generalization ability after model training, we selected 2,182 samples from approximately 4,000 otoendoscopic images as training samples and 475 samples as verification samples (Table 6). To select the deep CNN model, we tested the ResNet, SENet, and EfficientNet models with different layers. The optimal training results among the three models were selected. The average highest accuracies of ResNet101, SENet101, and EfficientNetB6 models were 78.32, 87.16, and 88.21%, respectively (Table 1). Considering the accuracy and speed of operation, we chose the EfficientNetB6 model to output the probability values of four types of otomycosis, impacted cerumen, external otitis, and the normal external auditory canal. After multiple iterative model training, the average accuracy of the overall validation sample was 92.42% (Table 3). The results suggest that the system could be used by doctors, or even patients, to better diagnose otomycosis.

We proposed a computer-aided decision system based on a deep learning model, which includes Java web application and image classification subsystems. The web application subsystem mainly provides a user-friendly page to collect images of consultation and display the calculation results. The image classification subsystem mainly uses a trained neural network model to perform end-to-end data reasoning. Finally, on

uploading a few otoendoscopic images, the system returns the classification results to the user as the category probability value.

We released a beta web application at http://175.178.230.136/. Guests can login with a username and password and upload otoendoscopic images up to four at a time for diagnoses. The image size should not exceed 2 MB, and the image should be in the JPEG format. The "picture identification" button should be clicked to obtain probability values of the four categories. The uploaded image should be clicked to enlarge, read, and confirm the identification. Figure 6 shows the screenshot of the identification results on the website.

4 Discussion

Otomycosis is mainly diagnosed based on clinical manifestations and mycological examination results. However, differences in the pathogenic fungi species directly affects the positive rate of direct microscopic examination results (Ali et al., 2018). By culturing isolated specimens, further morphological and molecular biological identifications can be carried out, and the pathogenic species can be identified (Merad et al., 2021). However, conventional identification methods are time-consuming and prone to cross-contamination of specimens, which often leads to failure of clinicians to select effective antifungal drugs at a timely and early stage, affecting the prognosis (Hagiwara et al., 2019). Otoendoscopy has the advantages of a broad field of vision, close observation, and less invasive injury and has been widely used in the diagnosis and treatment of outer and middle ear diseases in recent years (Ulku, 2017). To treat otomycosis, we usually remove the fungal focus under ear endoscopy and select the appropriate drug according to the fungal culture results. Therefore, otoendoscopy and fungal culture are required to diagnose otomycosis. In our study, otoendoscopic images of impacted cerumen, external otitis, and the normal external auditory canal were used simultaneously. The established image library accumulated sufficient data to diagnose otomycosis by deep learning of otoendoscopic images.

We developed a computer-aided support system to assist physicians in the diagnosis of otomycosis. To ensure a diagnostic accuracy comparable to that of ear, nose, and throat specialists and provide the best care to patients, the most appropriate feature extraction methods and learning models were selected. A neural network model with different layers was tested, and the best training results of the three models was selected considering the accuracy and operation speed. The performance of the EfficientnetB6 model was found to be the highest. The weighted mean values of the models in the four categories were obtained. The result of each model was multiplied by the weighted mean value of each model and classification to obtain the average accuracy of the total validation sample of the set classifier (Table 1, Table 2, and Table 3). Classical machine learning techniques, such as SVM, K-NN, and decision tree, provide high performances in classification tasks, particularly with reasonably sized datasets (Van Gestel et al., 2002). These techniques are easy to understand, simplifying model tuning and calibration. Other more complex models, such as CNNs, can be used to overcome the same challenges but must be trained with larger database to achieve comparable performance.

The binary classification method of still color images of the eardrum has been used to identify the normal ear and otitis media, with accuracy rates of 73.1 and 68.3%, respectively (Tran et al., 2018; Cai et al., 2021). In both cases, color information was used to train the learning models. However, color alone cannot be used to obtain an accurate classification. In a previous study, classifying cases of normal ear and otitis media (Shie et al., 2014), the color, texture, and geometric information were used to train support vector machines with an accuracy exceeding 88.1% achieved by previous authors. However, the system's specific ability to correctly identify healthy individuals was 79.9%. A study implemented a system to distinguish the normal eardrum, otitis media, and blocked ear canals with an accuracy of 86.8% (Myburgh et al., 2018). Whether these results were obtained through classification stages using validation or test sets remains unclear. The evaluation index of the deep CNN model selected in our study mainly depends on the performance of the set classifier, which is measured using the obfuscation matrix, precision, recall, and PR and ROC curves. The confusion matrix represents the number of instances corresponding to the predicted and actual classes (Table 4). To diagnose otomycosis, there were 187 images of otomycosis, 177, three, four, and three images of which were predicted to be of otomycosis, impacted cerumen, external otitis, and normal external auditory canal, respectively. Table 5 shows the statistics of the precision and recall. Figure 5 is drawn with the precision ratio as the vertical axis and the recall ratio as the horizontal axis to obtain the PR curve.

Previous studies on deep learning methods have distinguished between normal or abnormal conditions of the eardrum (Senaras et al., 2018). Two different deep learning architectures were used in this study, with an accuracy of 84.4%, a sensitivity of 85.9%, and a specificity of 82.8%. Another study proposed a diagnosis system based on deep CNN, with an average accuracy of 93.6% (Cha et al., 2019). The study divided ear diseases into five categories: invagination of the eardrum, perforation of the tympanum, tympanitis, external auditory canal tumors, and normal cases. However, the model performance, i.e., accuracy, depends on the number of images trained. If the number of images decreases, the performance degrades. A common limitation of deep learning approaches is the influence of database on the model (Zhang et al., 2020). In addition, a large database, particularly in otolaryngology, may be unavailable.

We introduced a novel beta web application, which is user-friendly. After several iterations of sample training, the average accuracy of the overall validation sample was 92.42%. Clinicians can login on the website and upload otoendoscopic images to accurately

diagnose otomycosis (Figure 6). In the future, doctors and patients would be able to upload images to their smartphones or other devices to obtain diagnoses by installing software. The web application has a learning function. Therefore, the program can learn the uploading of identification images for improved accuracy.

To create a database of images of otomycosis, we collected otoendoscopic images of the external auditory canal obtained in clinical practice. According to the clinical manifestations and otoendoscopic images, otomycosis was considered, and specimens from these patients were collected for fungal species identification. The diagnostic system was relatively easy to implement and could significantly impact primary healthcare. Most blurred and unqualified images are deleted, and a few blurred images can also be analyzed and processed by artificial intelligence algorithms, thus increasing the diagnostic or classification accuracy. Although we randomly selected a large sample of cases, all possible imaging presentations of otomycosis may not have been covered. Therefore, image selection will have a potential bias. However, to avoid bias and improve accuracy, our database includes images of different tympanic membranes and external auditory canal projections.

Finally, we only evaluated four conditions that were presented during endoscopy, while the diagnosis of external auditory canal mycosis should include the differential diagnosis of other rare diseases in clinical practice. Whether our method would show reduced or improved accuracy if more conditions are included remains unknown. Therefore, the characteristics of otoendoscopy should be evaluated in detail in future studies. Nevertheless, compared to previous studies, our study achieved greater accuracy in the diagnosis of otomycosis in a real clinical setting through conditions that had not been previously assessed.

In future studies, we aim to integrate other rare types of otomycosis, although their diagnosis is a challenge even for specialists. In addition, we would train deep CNNs for other learning models, which may allow the integration of more classifications to maintain high performance.

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 study protocol was approved by the Research Ethics Committee of Jingzhou Hospital Affiliated to Yangtze University (protocol number: 2021-093-01). Written informed consent was obtained from patients and caregivers of patients under 18 years of age. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CM and DP collection patients' samples, otoendoscopy images and analysis the experiment data. JW, YS, and AL designed, interpreted the experiment data and wrote the manuscript. JH and PW carried out the fungal pathogen examination. JW, AL, and CM revised the manuscript critically for important content. All authors read and approved the final manuscript.

Funding

This work was supported by Health Commission of Hubei Province scientific research project (Grant No. JW 2021M261 to YS), the Natural Science Foundation of Hubei Province (Grant No. 2019CFB567 to YS).

Acknowledgments

The authors would like to thank all the reviewers who participated in the review, as well as MJEditor (www.mjeditor.com) for providing English editing services during the preparation of this manuscript.

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

Ali, K., Hamed, M. A., Hassan, H., Esmail, A., and Sheneef, A. (2018). Identification of fungal pathogens in otomycosis and their drug sensitivity: our experience. *Int. Arch. Otorhinolaryngol.* 22, 400–403. doi:10.1055/s-0038-1626702

Cai, Y., Yu, J. G., Chen, Y., Liu, C., Xiao, L., Grais, M. E., et al. (2021). Investigating the use of a two-stage attention-aware convolutional neural network for the automated diagnosis of otitis media from tympanic membrane images: a prediction model development and validation study. *BMJ Open* 11, e041139. doi:10.1136/bmjopen-2020-041139

Cha, D., Pae, C., Seong, S. B., Choi, J. Y., and Park, H. J. (2019). Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine* 45, 606–614. doi:10.1016/j.ebiom.2019.06.050

Chougrad, H., Zouaki, H., and Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. *Comput. Methods Programs Biomed.* 157, 19–30. doi:10.1016/j.cmpb.2018.01.011

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi:10.1038/nature21056

Geras, K. J., Mann, R. M., and Moy, L. (2019). Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology* 293, 246–259. doi:10.1148/radiol.2019182627

Giger, M. L. (2018). Machine learning in medical imaging. *J. Am. Coll. Radiol.* 15, 512–520. doi:10.1016/j.jacr.2017.12.028

Godinez, W. J., Hossain, I., Lazic, S. E., Davies, J. W., and Zhang, X. (2017). A multi-scale convolutional neural network for phenotyping high-content cellular images. *Bioinformatics* 33, 2010–2019. doi:10.1093/bioinformatics/btx069

Hagiwara, S., Tamura, T., Satoh, K., Kamewada, H., Nakano, M., Shinden, S., et al. (2019). The molecular identification and antifungal susceptibilities of aspergillus species causing otomycosis in tochigi, Japan. *Mycopathologia* 184, 13–21. doi:10.1007/s11046-018-0299-9

Hapfelmeier, A., and Horsch, A. (2011). Image feature evaluation in two new mammography CAD prototypes. *Int. J. Comput. Assist. Radiol. Surg.* 6, 721–735. doi:10.1007/s11548-011-0549-5

Horsch, A., Hapfelmeier, A., and Elter, M. (2011). Needs assessment for next generation computer-aided mammography reference image databases and evaluation studies. *Int. J. Comput. Assist. Radiol. Surg.* 6, 749–767. doi:10.1007/s11548-011-0553-9

Kamali Sarwestani, Z., Hashemi, S. J., Rezaie, S., Gerami Shoar, M., Mahmoudi, S., Elahi, M.,, et al. (2018). Species identification and *in vitro* antifungal susceptibility testing of aspergillus section nigri strains isolated from otomycosis patients. *J. Mycol. Med.* 28, 279–284. doi:10.1016/j.mycmed.2018.02.003

Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., et al. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Med. Image Anal.* 35, 303–312. doi:10.1016/j.media.2016.07.007

Lee, J. H., Han, S. S., Kim, Y. H., Lee, C., and Kim, I. (2020). Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 129, 635–642. doi:10.1016/j.oooo.2019.11.007

Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., and Rubin, D. L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data* 4, 170177–177. doi:10.1038/sdata.2017.177

Li, Y., and He, L. (2019). Diagnosis and treatment of otomycosis in southern China. *Mycoses* 62, 1064–1068. doi:10.1111/myc.12979

Li, Y., Wang, H., Zhao, Y. P., Xu, Y. C., and Hsueh, P. R. (2020). Antifungal susceptibility of clinical isolates of 25 genetically confirmed aspergillus species collected from Taiwan and Mainland China. *J. Microbiol. Immunol. Infect.* 53, 125–132. doi:10.1016/j.jmii.2018.04.003

Merad, Y., Derrar, H., Belmokhtar, Z., and Belkacemi, M. (2021). Aspergillus genus and its various human superficial and cutaneous features. *Pathogens* 10, 643. doi:10.3390/pathogens10060643

Moses, D. A., Dawes, L., Sammut, C., and Zrimec, T. (2018). Peripheral bronchial identification on chest CT using unsupervised machine learning. *Int. J. Comput. Assist. Radiol. Surg.* 13, 1379–1395. doi:10.1007/s11548-018-1805-8

Myburgh, H. C., Jose, S., Swanepoel, D. W., and Laurent, C. (2018). Towards low cost automated smartphone- and cloud-based otitis media diagnosis. *Biomed. Signal Process. Control* 39, 34–52. doi:10.1016/j.bspc.2017.07.015

Senaras, C., Moberly, A. C., Teknos, T., Essig, G., Elmaraghy, C., Taj-Schaal, N., et al. (2018). Detection of eardrum abnormalities using ensemble deep learning approaches, *Proceeding Med. imaging 2018 Computer-Aided Diagnosis*, 10575. 105751A.

Shen, D., Wu, G., and Suk, H. I. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248. doi:10.1146/annurev-bioeng-071516-044442

Shie, C. K., Chang, H. T., Fan, F. C., Chen, C. J., Fang, T. Y., and Wang, P. C. (2014). A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 4655–4658. doi:10.1109/EMBC.2014.6944662

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* 35, 1285–1298. doi:10.1109/TMI.2016.2528162

Spasov, S. E., Passamonti, L., Duggento, A., Lio, P., and Toschi, N. (2018). A multi-modal convolutional neural network framework for the prediction of alzheimer's disease. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1271–1274. doi:10.1109/EMBC.2018.8512468

Talo, M., Yildirim, O., Baloglu, U. B., Aydin, G., and Acharya, U. R. (2019). Convolutional neural networks for multi-class brain disease detection using MRI images. *Comput. Med. Imaging Graph.* 78, 101673. doi:10.1016/j.compmedimag.2019.101673

Tan, L. K., Liew, Y. M., Lim, E., and McLaughlin, R. A. (2017). Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Med. Image Anal.* 39, 78–86. doi:10.1016/j.media.2017.04.002

Tran, T. T., Fang, T. Y., Pham, V. T., Lin, C., Wang, P. C., and Lo, M. T. (2018). Development of an automatic diagnostic algorithm for pediatric otitis media. *Otol. Neurotol.* 39, 1060–1065. doi:10.1097/MAO.0000000000001897

Ulku, C. H. (2017). Endoscopy-assisted ear Surgery for treatment of chronic otitis media with cholesteatoma, adhesion, or retraction pockets. *J. Craniofac. Surg.* 28 (4), 1017–1020. doi:10.1097/SCS.0000000000003671

Van Gestel, T., Suykens, J. A., Lanckriet, G., Lambrechts, A., De Moor, B., and Vandewalle, J. (2002). Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and kernel Fisher discriminant analysis. *Neural Comput.* 14, 1115–1147. doi:10.1162/089976602753633411

Varma, M., and Zisserman, A. (2009). A statistical approach to material classification using image patch exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 2032–2047. doi:10.1109/TPAMI.2008.182

Zhang, J., Xu, T., Shen, Z., Qiao, Y., and Zhang, Y. (2019). Fourier ptychographic microscopy reconstruction with multiscale deep residual network. *Opt. Express* 27, 8612–8625. doi:10.1364/OE.27.008612

Zhang, L., and Suganthan, P. N. (2017). Visual tracking with convolutional random vector functional link network. *IEEE Trans. Cybern.* 47, 3243–3253. doi:10.1109/TCYB.2016.2588526

Zhang, Q., Zhu, L., Bao, W., and Huang, D. S. (2020). Weakly-supervised convolutional neural network architecture for predicting protein-DNA binding. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17, 679–689. doi:10.1109/TCBB.2018.2864203

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 23 August 2022
DOI 10.3389/fmolb.2022.968458

Check for updates

OPEN ACCESS

EDITED BY
Teng Ma,
Capital Medical University, China

REVIEWED BY
Fengqiang Yang,
Tongji University, China
Hongyu Zhao,
Wake Forest University, United States

*CORRESPONDENCE
William C. Cho,
chocs@ha.org.hk
Dong Li,
lidong@tongji.edu.cn

†These authors have contributed equally to this work.

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 14 June 2022
ACCEPTED 02 August 2022
PUBLISHED 23 August 2022

CITATION
Chen C, Shang A, Gao Y, Huang J, Liu G, Cho WC and Li D (2022), PTBPs: An immunomodulatory-related prognostic biomarker in pan-cancer. *Front. Mol. Biosci.* 9:968458. doi: 10.3389/fmolb.2022.968458

COPYRIGHT
© 2022 Chen, Shang, Gao, Huang, Liu, Cho 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.

PTBPs: An immunomodulatory-related prognostic biomarker in pan-cancer

Chen Chen[1†], Anquan Shang[1†], Yuting Gao[1], Jingjuan Huang[1], Gege Liu[1], William C. Cho[2*] and Dong Li[1*]

[1]Department of Laboratory Medicine, Shanghai Tongji Hospital, School of Medicine, Tongji University, Shanghai, China, [2]Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong, Hong Kong SAR, China

Background: The polypyrimidine tract-binding protein (PTBP) nuclear ribonucleoprotein family of proteins, including PTBP1, PTBP2 and PTBP3, regulate the process of cell proliferation, differentiation, apoptosis and carcinogenesis. PTBPs exhibit oncogenic effects in certain tumors. However, the role of PTBPs in pan-cancer remains unclear. Our study examined the clinical significance and mechanism of PTBPs in pan-cancer.

Methods: We compared the expression of *PTBPs* in paired and unpaired tissue samples from the Cancer Genome Atlas (TCGA) database. Univariate and multivariate Cox regression, Kaplan–Meier curves, and time-dependent receiver operating characteristic (ROC) curves were used to assess the prognostic significance of *PTBPs* in pan-cancer. The cBioPortal database also identified genomic abnormalities in *PTBPs*. TISIDB, TCGA, and Cellminer were used to investigate the relationship between *PTBP* expression and immune subtypes, immune checkpoint (ICP) genes, tumor mutational burden (TMB), microsatellite instability (MSI), tumor-infiltrating immune cells, and chemosensitivity. cBioPortal was used to search for *PTBP* co-expressing genes in pan-cancer, and GO and KEGG enrichment analyses were performed to search for *PTBP*-related signaling pathways.

Results: *PTBPs* were shown to be widely upregulated in human tumor tissues. *PTBP1* showed good prognostic value in ACC, KIRP, and LGG; *PTBP2* in ACC and KICH; and *PTBP3* in ACC, LGG, and PAAD, with AUC >0.7. *PTBPs* were differentially expressed in tumor immune subtypes and had a strong correlation with tumor-infiltrating lymphocytes (TILs) in the tumor microenvironment (TME). In addition, *PTBP* expressions were related to ICP, TMB, and MSI, suggesting that these three PTBPs may be potential tumor immunotherapeutic targets and predict the efficacy of immunotherapy. Enrichment analysis of co-expressed genes of *PTBPs* showed that they may be involved in alternative splicing, cell cycle, cellular senescence, and protein modification.

Conclusion: PTBPs are involved in the malignant progression of tumors. *PTBP1*, *PTBP2* and *PTBP3* may be potential biomarkers for prognosis and immunotherapy in pan-cancer and may be novel immunotherapeutic targets.

KEYWORDS

biomarkers, immunotherapy, pan-cancer, polypyrimidine tract-binding proteins, prognosis

Introduction

Cancer is a life-threatening disease to humans worldwide. The incidence and mortality rates of various cancers have been increasing year by year, and lung cancer, colorectal cancer, liver cancer and gastric cancer have the highest mortality rates (Sung et al., 2021). The field of precision medicine is advancing through new developments in technology and medicine, but the present state of practice is far from ideal (König et al., 2017). Therefore, the identification of tumor-related diagnostic, prognostic, and therapeutic biomarkers is a research hotspot.

Polypyrimidine tract-binding proteins (PTBPs) are important RNA-binding proteins (RBPs) which influence cell growth and development by regulating mRNA stability, translation and alternative splicing (Singh et al., 1995; Mickleburgh et al., 2014). The PTBP family consists of PTBP1, PTBP2 and PTBP3, and these proteins show similarities and differences in expression, structure, and biological function (Oberstrass et al., 2005; Tan et al., 2015). Studies have shown that up-regulation of PTBP1 is associated with the poor prognosis and disease progression in non-muscle-invasive bladder cancer. Therefore, PTBP1 may become a possible outcome-predictor for bladder cancer (Bielli et al., 2018). PTBP2 was also shown to stimulate the proliferation, migration, and metastasis of colorectal cancer cells (Ji et al., 2014). Previous studies demonstrated that PTBP3 enhances the invasion and metastasis of breast cancer and regulates the expression of drug resistance proteins in gastric cancer, suggesting that PTBP3 may serve as a potential novel therapeutic target for gastric cancer (Liang et al., 2017; Hou et al., 2018; Liang et al., 2020). These studies have indicated that PTBPs may function in cancer. However, research has been restricted to a small number of tumor types, and the function of PTBPs in pan-cancer has not been examined.

A growing body of evidence has revealed the close relationship between the tumor microenvironment (TME) and the effectiveness of immunotherapy (Donlon et al., 2021; Newnes et al., 2021). Immune checkpoint (ICP) inhibitors, including PD-1, CTLA4, LAG3, and TIM-3, have potent tumor suppressor effects and can interfere with immune escape; these inhibitors are currently the first-line treatment options for multiple malignancies (Ribas and Wolchok, 2018; Tu et al., 2020). Tumor infiltrating immune cells in TME, such as macrophages, neutrophils, T cells, Treg cells, T helpers, and NK cells, can affect the immunological features of malignancies. Unfortunately, this tumor heterogeneity among individuals influences the efficacy of clinical immunotherapy (Keenan et al., 2019). Thus far, precision medicine has not completely manifested in human tumors. Researchers agree on the need to explore better treatment targets.

In this study, we evaluated the clinical importance and prognostic usefulness of *PTBPs* in pan-cancer. We used the Cancer Genome Atlas (TCGA) to examine the expression levels of *PTBP1*, *PTBP2*, and *PTBP3* in normal and tumor tissues, and cBioPortal was used to examine the genomic alterations. We also examined the link between *PTBP* expression and tumor immune subtype, tumor-infiltrating lymphocytes (TILs), ICP, tumor mutational burden (TMB), microsatellite instability (MSI), and chemosensitivity using multiple databases. Finally, we constructed a PTBP-interacting protein network and performed enrichment analysis of co-expressed genes. Our findings have demonstrated the prognostic value of *PTBPs* in pan-cancer. PTBPs may have excellent potential to be therapeutic targets and predict the efficacy of immunotherapy. We also predicted the molecular mechanisms and biological signaling pathways of *PTBPs* using databases and experimental data.

Materials and methods

Study overview

A total of 30 tumor types were studied in this article. A schematic flow chart of our research is shown in Supplementary Figure S1.

Difference and correlation analysis of *PTBP1*, *PTBP2*, and *PTBP3* expression

Gene expression, clinical data, and survival information for 30 different types of tumors were downloaded from TCGA database (https://portal.gdc.cancer.gov), and the RNA-seq data in level 3 HTSeq-FPKM format were log2-transformed. The Mann–Whitney *U* test was used to analyze the differences in the expression levels of *PTBP1*, *PTBP2*, and *PTBP3* in unpaired tissue samples, and the Wilcoxon signed rank test was used for paired samples.

To analyze the expression correlation between *PTBP1*, *PTBP2*, and *PTBP3* genes, we excluded the tumor types with less than three normal samples and then log2-transformed the ratio of the mean expression of *PTBPs* in tumors and normal samples of the remaining 21 tumor types. All details are shown in Supplementary Table S2.

Copy number alterations and mutations

Copy number alterations and mutations of *PTBP* genes were analyzed using the online database cBioPortal (http://www.cbioportal.org).

Cox regression analysis, Kaplan–Meier curve, time-dependent receiver operating characteristic curve and prognostic nomogram

First, univariate Cox regression analysis was performed on the expressions of *PTBP1*, *PTBP2*, and *PTBP3* in pan-cancer. Factors with $p \leq 0.05$ were included in multivariate Cox regression analysis and displayed in a forest plot. We used Cox regression models to predict survival and plotted Kaplan–Meier (KM) curves. The accuracy of the model in predicting prognosis at a specific time was tested by a time-dependent receiver operating characteristic (ROC) curve, and the prognostic value of *PTBPs* in pan-cancer was determined. Taking the tumor "ACC" as an example, we constructed prognostic nomograms using *PTBP* expressions and pathological stage, and the accuracy of the nomogram was evaluated by a calibration curve. We examined overall survival (OS) in our analyses.

Analysis of *PTBP* expression and tumor immune subtypes

We analyzed the associations between *PTBP* expression and immune subtypes in human cancers using TISIDB, an online integrated website (http://cis.hku.hk/TISIDB/index.php). $p \leq 0.05$ was considered statistically significant.

Relationship between *PTBP* expression and tumor-infiltrating lymphocytes in pan-cancer

Using the ssGSEA algorithm to calculate the score of tumor-infiltrated immune cells in TCGA database, we selected 24 TILs to evaluate the relationship between *PTBP* expressions and tumor-infiltrating lymphocytes (TILs) in pan-cancer. The TILs included activated DCs (aDCs), B cells, CD8 T cells, cytotoxic cells, DC, eosinophils, immature DCs (iDCs), macrophages, mast cells, neutrophils, NK CD56bright cells, NK CD56dim cells, NK cells, plasmacytoid DCs (pDCs), T cells, T helper cells, T central memory (Tcm), T effector memory (Tem), T follicular helper (Tfh), T gamma delta (Tgd), Th1 cells, Th17 cells, Th2 cells, and Tregs (Bindea et al., 2013). Neutrophils and macrophages were selected for detailed display.

Correlation analysis of *PTBP* expressions with immune checkpoint, tumor mutational burden, and microsatellite instability

Spearman correlation analysis was performed to examine the relation of the expression levels of *PTBPs* and four immune checkpoint (ICP) genes (PD-1, CTLA4, LAG3, and TIM-3 genes). We used Spearman correlation analysis to determine the correlation of *PTBP* expressions with tumor mutational burden (TMB) and MSI in human cancers. RNA-seq data and clinical information were obtained from TCGA database. The TMB data and microsatellite instability (MSI) data were derived from the studies of Thorsson et al. (2018) and Bonneville et al. (2017), respectively. $p \leq 0.05$ was considered statistically significant.

Correlation analysis with drug susceptibility

Transcriptome data (RNA: RNA-seq) and drug sensitivity data (compound activity: DTP NCI-60) were downloaded from Cellminer (https://discover.nci.nih.gov/cellminer); these data are derived from the same 60 samples. Only FDA approved samples were analyzed. A positive correlation indicated that the higher the expression of *PTBP*, the more sensitive that the cells were to the drugs.

Protein-protein interaction network analysis

We performed the analysis of Protein-Protein Interaction (PPI) networks using the STRING website (https://string-db.org/). The parameters for finding the interacting proteins of PTBPs were set as follows: minimum required interaction score ["Medium confidence (0.400)"], meaning of network edges ("evidence") and active interaction sources ("experiments"). The parameters for finding the interaction relationship among PTBP1, PTBP2 and PTBP3 were set as follows: minimum required interaction score ["Medium confidence (0.400)"], meaning of network edges ("evidence"), active interaction sources ("textmining, experiments, databases, co-expression,

co-occurrence"), and max number of interactors to show ("no more than 10 interactors").

Co-expressed genes of *PTBPs* in pan-cancer

We downloaded the dataset "Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020)" on cBioPortal and screened several mRNAs co-expressed with *PTBPs* using the absolute value of Spearman's correlation coefficient ≥0.4 and $p < 0.05$. Data were obtained from 991 samples.

GO enrichment and KEGG pathway analysis

We performed ID conversion on 926 *PTBP1*-related mRNAs, 657 *PTBP2*-related mRNAs, and 874 *PTBP3*-related mRNAs obtained in the previous step and analyzed their functions by GO and KEGG enrichment analysis. After correcting the *p*-value by the BH method, items with p. adjust ≤0.05 were selected for partial visualization.

Statistical analysis

R (version 3.6.3) was used for statistical analysis and visualization. The following R packages were used in this study: ggplot2 package [version 3.3.3], ggpubr package [version 0.1.4] for basic drawing, limma package [version 3.28.14] for differential analysis, survminer package [version 0.4.9], survival package [version 3.2–10] for statistical analysis of survival data, timeROC package [version 0.4] for ROC curve analysis, rms package [version 6.2–0] for building nomograms, impute package [version 1.68.0] for processing the missing value, GSVA package [version 1.34.0] for immune infiltration analysis, org. Hs.eg.db package [version 3.10.0] for id conversion, and the clusterProfiler package [version 3.14.3] for enrichment analysis.

Results

Expression levels of *PTBPs* in tissues of pan-cancer

We analyzed the expression levels of *PTBPs* in tumor tissues and normal/adjacent tissues from 30 cancer types, and the abbreviations of tumor types are shown in Supplementary Table S1. Paired differential expression analysis was then performed for 18 cancer types with more than three normal samples. Among *PTBPs*, *PTBP1* showed the highest RNA expression level in tumor tissues, followed by *PTBP3*; the expression level of *PTBP2* was the lowest among the three genes. *PTBP1* was upregulated in tumor tissues compared with normal tissues in BLCA, BRCA, CESC, CHOL, COAD, ESCA, GBM, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, READ, STAD, and UCEC. *PTBP2* was upregulated in CHOL, HNSC, KIRC, LIHC, LUAD, and LUSC and downregulated in BLCA, BRCA, CESC, GBM, KICH, PCPG, PRAD, READ, THCA, and UCEC. *PTBP3* was upregulated in BLCA, BRCA, CESC, CHOL, COAD, ESCA, GBM, HNSC, LIHC, LUAD, LUSC, READ, STAD, and UCEC and downregulated in KICH, KIRC, KIRP, and THCA (Figures 1A–C). The results of differential expression analysis for paired samples are shown in Figures 1D–F. Furthermore, we predicted the structures of these three proteins based on AlphaFold (Jumper et al., 2021; Varadi et al., 2022) (Supplementary Figures S2–S4), among which, PTBP1 and PTBP2 were highly similar in structure.

Expression correlation among *PTBP1, PTBP2,* and *PTBP3*

Expression correlation analysis was performed on 21 cancer types with more than three normal samples. As shown in the circular heatmap in Figure 1G, *PTBP1* and *PTBP3* are expressed similarly, while *PTBP2* and *PTBP1/3* have the opposite expression in multiple cancers.

Genetic alterations of *PTBPs*

We next used the public dataset "Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020)" from cBioPortal to examine copy number alterations and mutations. The dataset includes a total of 2565 patients, and information on these three genes was available in 174 patients. The copy number alterations and mutation data of *PTBPs* in pan-cancer are shown in Figure 2A. *PTBP1* was the most altered among the *PTBPs*, with the main genetic alteration types being amplifications and deep deletions. *PTBP1* was frequently altered in colorectal cancer, pancreatic cancer, and ovarian cancer; *PTBP2* was frequently altered in non-small cell lung cancer, lung cancer, and ovarian cancer; and *PTBP3* was frequently altered in pancreatic cancer, soft tissue sarcoma, and colorectal cancer (Figures 2B–D).

PTBPs are clinically significant tumor-associated factors

We next analyzed the relationship between *PTBP* expression and stage, grade, or other clinical features in

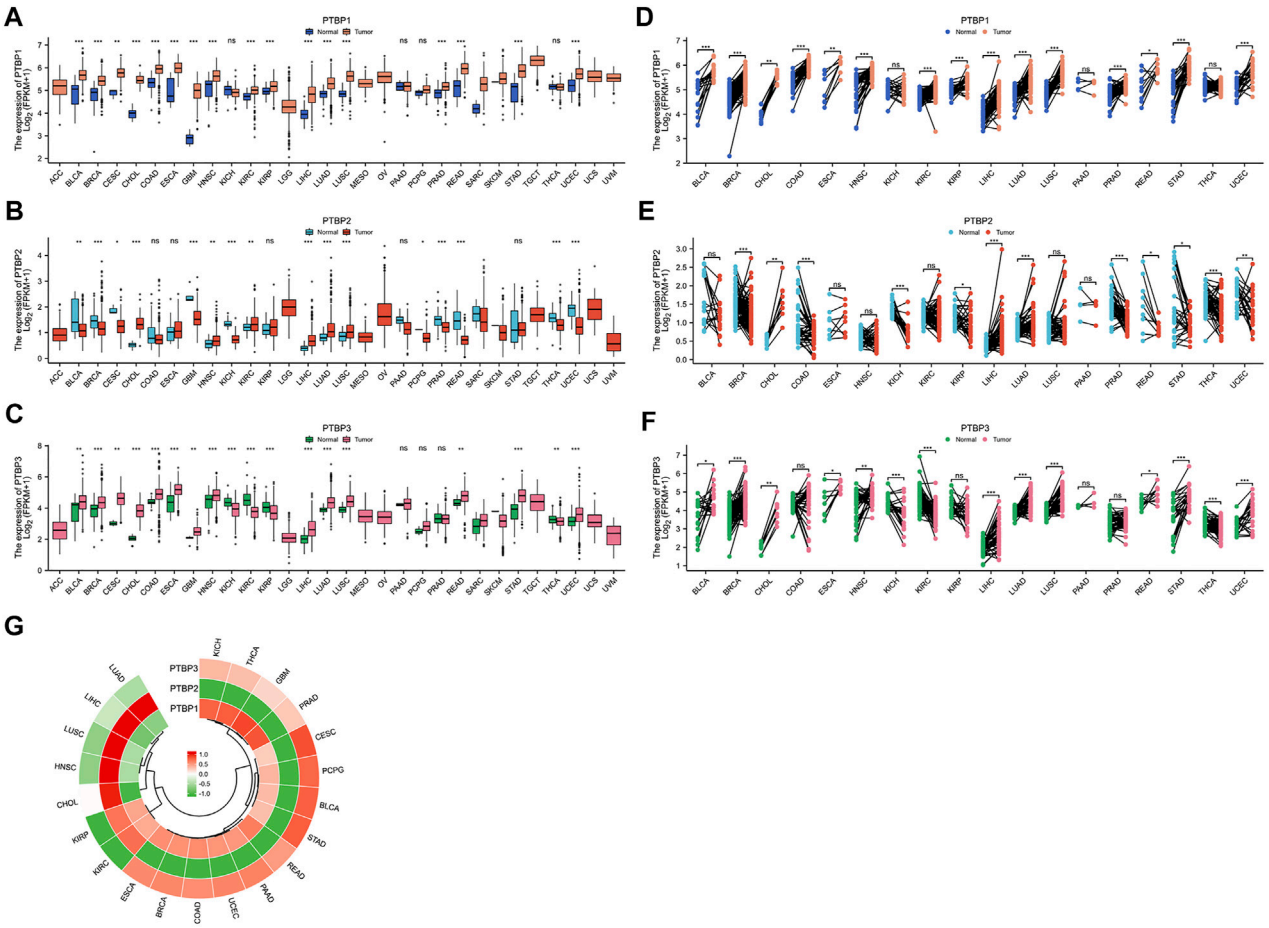


FIGURE 1
The expression levels of *PTBP* in pan-cancer tissues and the expression correlation among the three *PTBP* genes. RNA-seq data were obtained from the TCGA database. **(A–C)** Differential expression of *PTBPs* in unpaired samples of 30 tumor types. **(D–F)** Differential expression of *PTBPs* in paired samples of 30 tumor types (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns: not significant). **(G)** Correlation analysis of the expressions of *PTBP1*, *PTBP2*, and *PTBP3* in various tumor types.

pan-cancer and drew violin plots using data from TCGA (Figure 3). The expression of *PTBP* has a strong correlation with the clinical characteristics of various tumors, indicating that *PTBPs* are related to the occurrence and development of tumors. For example, the expression of *PTBP3* was significantly higher in high-grade BLCA than that in low-grade BLCA ($p = 0.002$). *PTBP2* expression in IDH-mutant GBM was significantly higher than that in IDH-wildtype GBM ($p < 0.001$). The expression of *PTBP1* was related to the pathological stage of ACC: *PTBP1* expression in stage IV ACC was significantly higher than that in stage I ACC ($p = 0.004$). Multivariate Cox regression analysis showed that the HR of *PTBP1* in ACC was 3.97 ($p = 0.01$) (Figure 4A). Patients with high expression of *PTBP1* had a lower probability of survival (Figure 4D), indicating that *PTBP1* is an independent risk factor for ACC and has the potential to be a prognostic indicator.

Prognostic value of *PTBPs* in pan-cancer

We downloaded RNA-seq data and corresponding clinical information of tumor tissues from 30 cancer types in TCGA database. Univariate and multivariate Cox regression analyses were performed on the OS data to analyze the hazard ratio (HR), 95% confidence interval (95%CI), and p value of *PTBPs* in pan-cancer (Table 1). Through multivariate Cox regression analysis, we found that *PTBP1* was significantly associated with poor prognosis in ACC, KIRP, LGG, LUAD, MESO, PRAD, and SKCM (HR > 1, $p < 0.05$). *PTBP2* is a risk factor for ACC, KICH, LIHC, and UCEC (HR > 1, $p < 0.05$) and a protective factor for OV, SKCM, and UCS (HR < 1, $p < 0.05$). *PTBP3* was significantly associated with poor prognosis in ACC, LGG, PAAD, and PCPG (HR > 1, $p < 0.05$), but predicted better prognosis in KICH and KIRC (HR < 1, $p < 0.05$). The above-listed tumors were chosen to make the forest plots (Figures 4A–C).

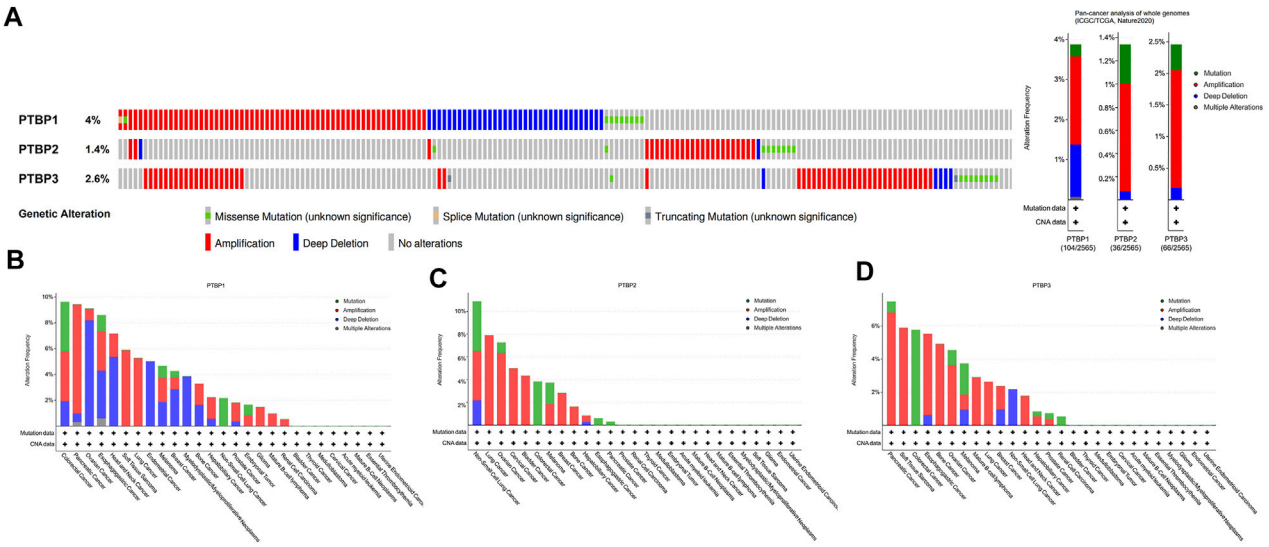


FIGURE 2
Characterization of genetic alterations in *PTBPs*. **(A)** General profile of genetic alterations in *PTBPs* in the pan-cancer dataset from cBioPortal. **(B–D)** Genetic alterations of *PTBPs* in specific tumor types, in descending order of alteration frequency.

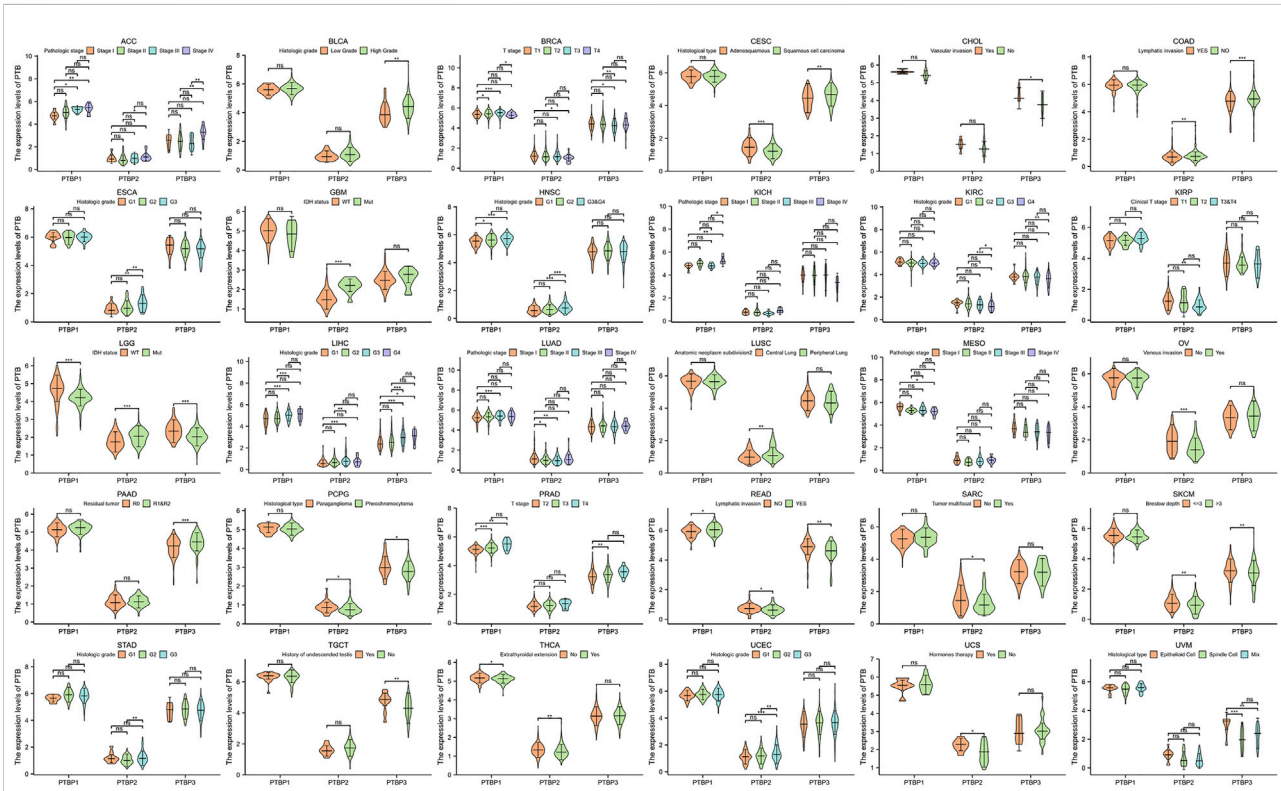


FIGURE 3
Expression levels of *PTBPs* in clinical parameters of interest (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ns: not significant).

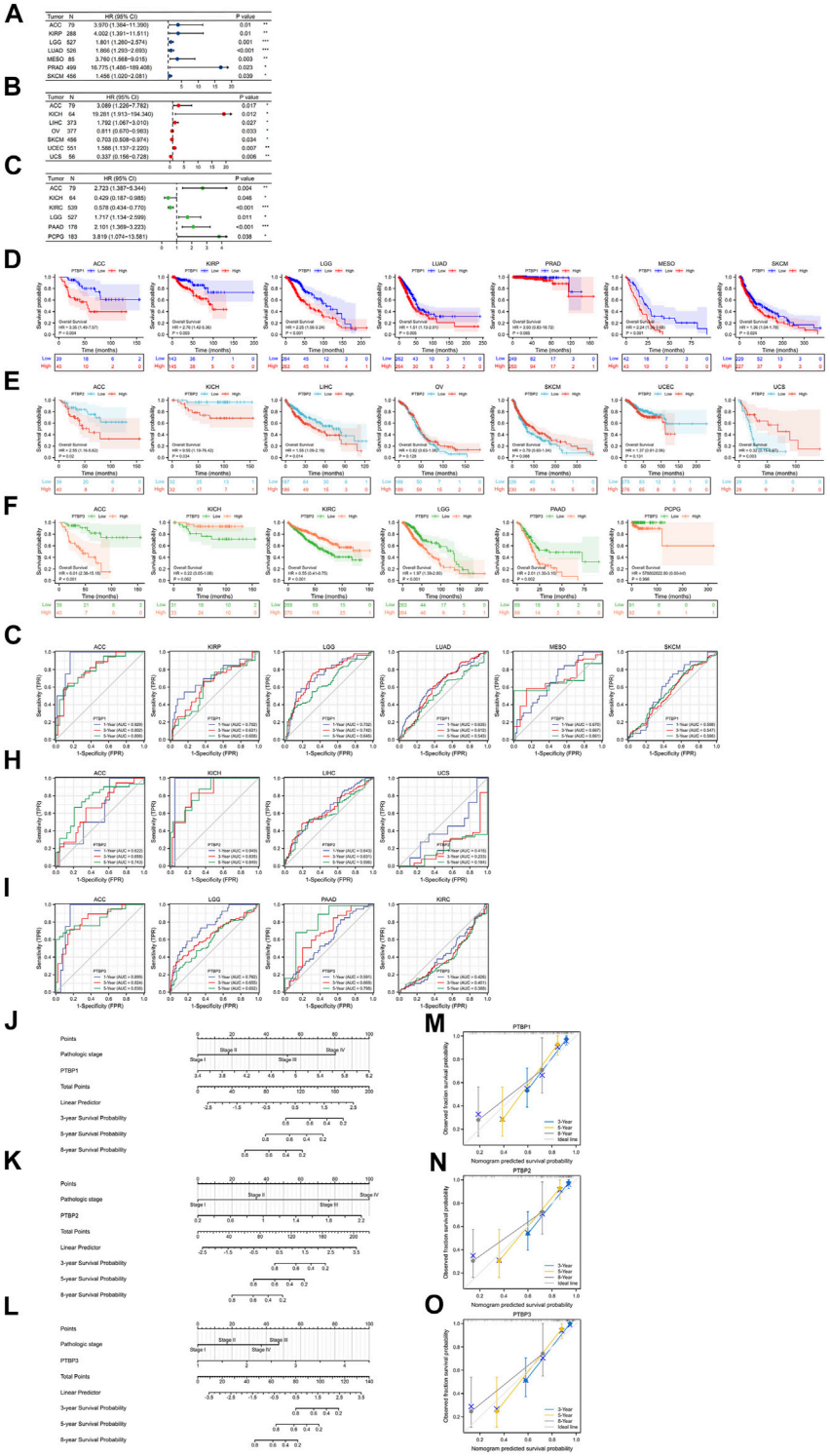


FIGURE 4
Evaluation of the prognostic value of *PTBPs* in pan-cancer. The multivariate Cox regression analysis results of **(A)** *PTBP1*, **(B)** *PTBP2*, and **(C)** *PTBP3* were visualized and presented as forest plots (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(D–F)** OS-KM survival curves of *PTBPs* in multiple cancers. **(G–I)** Time-dependent ROC curves of *PTBPs* to evaluate the utility of *PTBPs* as prognostic markers in selected tumor types. **(J–O)** Prognostic nomograms of *PTBP* expression combined with pathological stage in ACC. **(M–O)** The calibration curves for nomograms.

TABLE 1 Univariate and multivariate Cox regression analysis of PTBP expressions and overall survival in pan-cancer.

Cancer (OS)	n	PTBP1 univariate analysis			PTBP1 multivariate analysis			PTBP2 univariate analysis			PTBP2 multivariate analysis			PTBP3 univariate analysis			PTBP3 multivariate analysis		
		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value	
ACC	79	6.748 (2.433–18.713)	<0.001	***	3.970 (1.384–11.390)	0.010	**	4.346 (1.859–10.159)	<0.001	***	3.089 (1.226–7.782)	0.017	*	4.343 (2.376–7.940)	<0.001	***	2.723 (1.387–5.344)	0.004	**
BLCA	413	0.849 (0.551–1.308)	0.457					0.828 (0.554–1.237)	0.357					0.892 (0.711–1.120)	0.325				
BRCA	1082	0.911 (0.532–1.560)	0.733					0.787 (0.529–1.171)	0.238					1.226 (0.924–1.628)	0.158				
CESC	306	1.139 (0.475–2.732)	0.771					0.640 (0.338–1.215)	0.173					1.012 (0.641–1.598)	0.96				
CHOL	36	0.450 (0.066–3.089)	0.416					0.711 (0.269–1.876)	0.491					0.464 (0.163–1.322)	0.151				
COAD	477	1.239 (0.747–2.056)	0.406					0.906 (0.537–1.529)	0.711					0.890 (0.682–1.162)	0.392				
ESCA	162	1.121 (0.496–2.536)	0.783					0.931 (0.514–1.687)	0.814					1.208 (0.729–2.001)	0.463				
GBM	168	1.061 (0.765–1.472)	0.723					0.684 (0.468–1.002)	0.051					1.012 (0.672–1.523)	0.954				
HNSC	501	0.988 (0.670–1.456)	0.950					0.640 (0.406–1.007)	0.054					1.054 (0.823–1.350)	0.675				
KICH	64	38.073 (2.640–549.096)	0.008	**	4.222 (0.280–63.719)	0.298	ns	44.174 (6.571–296.980)	<0.001	***	19.281 (1.913–194.340)	0.012	*	0.425 (0.191–0.945)	0.036	*	0.429 (0.187–0.985)	0.046	*
KIRC	539	1.267 (0.769–2.086)	0.353					1.232 (0.832–1.825)	0.298					0.578 (0.434–0.770)	<0.001	***	0.578 (0.434–0.770)	<0.001	***
KIRP	288	4.002 (1.391–11.511)	0.010	**	4.002 (1.391–11.511)	0.010	**	0.692 (0.342–1.402)	0.306					1.318 (0.837–2.076)	0.233				
LGG	527	2.424 (1.802–3.262)	<0.001	***	1.801 (1.260–2.574)	0.001	***	0.684 (0.442–1.060)	0.089					2.702 (1.900–3.844)	<0.001	***	1.717 (1.134–2.599)	0.011	*
LIHC	373	1.850 (1.256–2.725)	0.002	**	1.663 (0.994–2.782)	0.053	ns	2.087 (1.370–3.179)	<0.001	***	1.792 (1.067–3.010)	0.027	*	1.382 (1.068–1.789)	0.014	*	0.909 (0.617–1.338)	0.627	ns
LUAD	526	1.866 (1.293–2.693)	<0.001	***	1.866 (1.293–2.693)	<0.001	***	0.792 (0.586–1.072)	0.131					1.184 (0.923–1.517)	0.183				
LUSC	496	1.030 (0.728–1.456)	0.868					0.934 (0.666–1.311)	0.695					1.238 (0.944–1.625)	0.123				
MESO	85	3.760 (1.568–9.015)	0.003	**	3.760 (1.568–9.015)	0.003	**	0.964 (0.448–2.072)	0.925					1.377 (0.916–2.070)	0.124				
OV	377		0.723						0.033	*		0.033	*		0.205				

(Continued on following page)

10.3389/fmolb.2022.968458

TABLE 1 (*Continued*) Univariate and multivariate Cox regression analysis of PTBP expressions and overall survival in pan-cancer.

Cancer (OS)	n	PTBP1 univariate analysis			PTBP1 multivariate analysis			PTBP2 univariate analysis			PTBP2 multivariate analysis			PTBP3 univariate analysis			PTBP3 multivariate analysis		
		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value		HR (95%CI)	*p* value	
		0.948 (0.707–1.272)						0.811 (0.670–0.983)			0.811 (0.670–0.983)			1.160 (0.922–1.458)					
PAAD	178	1.082 (0.561–2.087)	0.814					0.600 (0.326–1.106)	0.102					2.101 (1.369–3.223)	<0.001	***	2.101 (1.369–3.223)	<0.001	***
PCPG	183	7.151 (0.269–190.171)	0.240					0.053 (0.001–2.006)	0.113					3.819 (1.074–13.581)	0.038	*	3.819 (1.074–13.581)	0.038	*
PRAD	499	23.975 (2.154–266.880)	0.010	**	16.775 (1.486–189.408)	0.023	*	5.099 (0.489–53.212)	0.173					3.753 (0.941–14.970)	0.061				
READ	166	0.448 (0.162–1.238)	0.122					1.281 (0.287–5.718)	0.745					0.629 (0.335–1.180)	0.149				
SARC	263	2.586 (1.552–4.310)	<0.001	***	3.207 (0.635–16.210)	0.159	ns	0.918 (0.690–1.220)	0.554					1.278 (0.905–1.806)	0.163				
SKCM	456	1.491 (1.048–2.122)	0.026	*	1.456 (1.020–2.081)	0.039	*	0.685 (0.494–0.950)	0.023	*	0.703 (0.508–0.974)	0.034	*	0.892 (0.722–1.103)	0.292				
STAD	370	0.742 (0.532–1.036)	0.079					1.193 (0.838–1.699)	0.326					0.856 (0.663–1.105)	0.232				
THCA	510	1.098 (0.088–13.772)	0.942					2.768 (0.717–10.688)	0.140					1.960 (0.615–6.249)	0.255				
TGCT	139	0.476 (0.031–7.289)	0.594					0.364 (0.036–3.691)	0.393					3.799 (0.587–24.563)	0.161				
UCEC	551	0.716 (0.425–1.206)	0.209					1.588 (1.137–2.220)	0.007	**	1.588 (1.137–2.220)	0.007	**	1.008 (0.777–1.307)	0.955				
UCS	56	1.653 (0.665–4.110)	0.280					0.337 (0.156–0.728)	0.006	**	0.337 (0.156–0.728)	0.006	**	1.335 (0.786–2.268)	0.285				
UVM	80	1.319 (0.341–5.098)	0.688					0.891 (0.356–2.231)	0.805					1.822 (0.967–3.434)	0.064				

10.3389/fmolb.2022.968458

We divided samples into high and low expression groups using the median value of *PTBP* expression, predicted survival possibility, and plotted OS-KM curves. Except for PRAD ($p > 0.05$), the survival time of patients with ACC, KIRP, LGG, LUAD, MESO, and SKCM was shorter when *PTBP1* was highly expressed, with statistical significance (HR > 1, $p \leq 0.05$) (Figure 4D). Patients with ACC, KICH, and LIHC with high *PTBP2* expression had a shorter OS (HR > 1, $p \leq 0.05$), while patients with UCS with high *PTBP2* expression had a longer survival time (HR = 0.32, $p = 0.003$). No significant differences in OS were observed in OV, SKCM, and UCEC ($p > 0.05$) (Figure 4E). When *PTBP3* was highly expressed, patients with ACC, LGG, and PAAD had lower survival probability and poorer prognosis (HR > 1, $p \leq 0.05$), while patients with KIRC had a longer OS and better prognosis (HR = 0.55, $p < 0.001$); there was no significant differences observed in patients with KICH and PCPG ($p > 0.05$) (Figure 4F).

By comprehensively analyzing the results of multivariate Cox regression and the OS-KM curve, we concluded that *PTBP1* is a risk factor for ACC, KIRP, LGG, LUAD, MESO, and SKCM, and high expression of *PTBP1* predicts a shorter survival time. *PTBP2* is a risk factor in ACC, KICH, and LIHC and a protective factor in UCS. *PTBP3* is a risk factor for ACC, LGG, and PAAD but a protective factor for KIRC.

Next, we assessed the prognostic value of the three genes in the above tumors. We analyzed the predictive ability of *PTBP* genes for prognosis at 1, 3, and 5 years by time-dependent ROC curves to confirm the accuracy of these candidate markers (Supplementary Table S3). Our results indicated that *PTBP1* may serve as a prognostic biomarker for ACC, KIRP, and LGG at the three time points according to the criterion of AUC >0.7 (Figure 4G). *PTBP2* showed good prognostic value in ACC and KICH (Figure 4H), and *PTBP3* showed good prognostic value in ACC, LGG, and PAAD (Figure 4I), indicating these *PTBPs* may function as prognostic biomarkers in these tumors.

Our results showed that all three *PTBPs* were associated with poor prognosis in ACC, and therefore we determined the prognostic nomogram of ACC. Pathologic stage and *PTBP* expression were included in Cox regression analysis to establish prognostic nomograms. A vertical line was drawn to connect corresponding points and calculate the total score to estimate the 3-, 5-, and 8-years survival probability of ACC patients (Figures 4J–L). Calibration curves used to observe the predictive effect of the nomogram are shown in Figures 4M–O.

The expression of *PTBPs* in tumor immune subtypes

In 2018, Scientists performed an extensive immune genomic analysis of 33 cancer types (Thorsson et al., 2018). Six immune subtypes, including C1 (wound healing), C2 (IFN-gamma dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet), and C6 (TGF-β dominant), were identified by macrophage and lymphocyte markers, the ratio of Th1 cells to Th2 cells, and immune regulatory genes. This tumor heterogeneity leads to suboptimal outcomes of immunotherapy in the clinic.

We investigated the expression levels of *PTBPs* in different tumor immune subtypes by TISIDB. The results showed that *PTBP1* expression was associated with tumor immune subtypes of BLCA, BRCA, COAD, ESCA, GBM, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, PRAD, SARC, SKCM, STAD, TGCT, and UCEC (Figure 5A). *PTBP2* expression correlated with tumor immune subtypes of BLCA, BRCA, LGG, LUAD, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT, THCA, and UCEC (Figure 5B). *PTBP3* expression correlated with tumor immune subtypes of BLCA, BRCA, ESCA, KIRC, KIRP, LGG, LIHC, LUAD, LUSC, OV, SARC, SKCM, STAD, TGCT, and UCEC (Figure 5C). The expression of *PTBPs* in immune subtypes of other cancers is shown in Supplementary Figures S5–S7.

PTBP expression and immune infiltrating cells in the tumor microenvironment

The Spearman correlations between *PTBPs* and TILs in various tumor types were further investigated. We found strong positive correlations of *PTBP1* with Th2 cells (Figure 6A), *PTBP2* with T helper cells and Tcm (Figure 6B), and *PTBP3* with T helper cells, Tcm, and Th2 cells (Figure 6C) in most tumor types. Overall, the expression of *PTBPs* was highly correlated with the number of TILs in the TME. These results suggest that PTBPs may have a regulatory effect on the tumor microenvironment (TME) (Geng et al., 2021).

We also examined *PTBP* expression with macrophages and neutrophils (Figures 6D–F). Most tumor types had an infiltration of macrophages and neutrophils, and this was inversely linked with the expression of *PTBP1* and *PTBP2*. However, in many tumor types, *PTBP3* levels were favorably linked with macrophage and neutrophil counts. These details are shown in Supplementary Table S4.

Correlation analysis of *PTBPs* and immune checkpoint

Studies have shown that the immune checkpoint (ICP) genes have a great influence on the efficacy of immunotherapy. PD-1, CTLA4, LAG3, and TIM-3 are four ICPs that are frequently examined in the clinic, and inhibitors targeting these factors have shown potent tumor-killing effects in a variety of tumors (Sun et al., 2021; Yang et al., 2021; Gaikwad et al., 2022; Tian et al., 2022). To

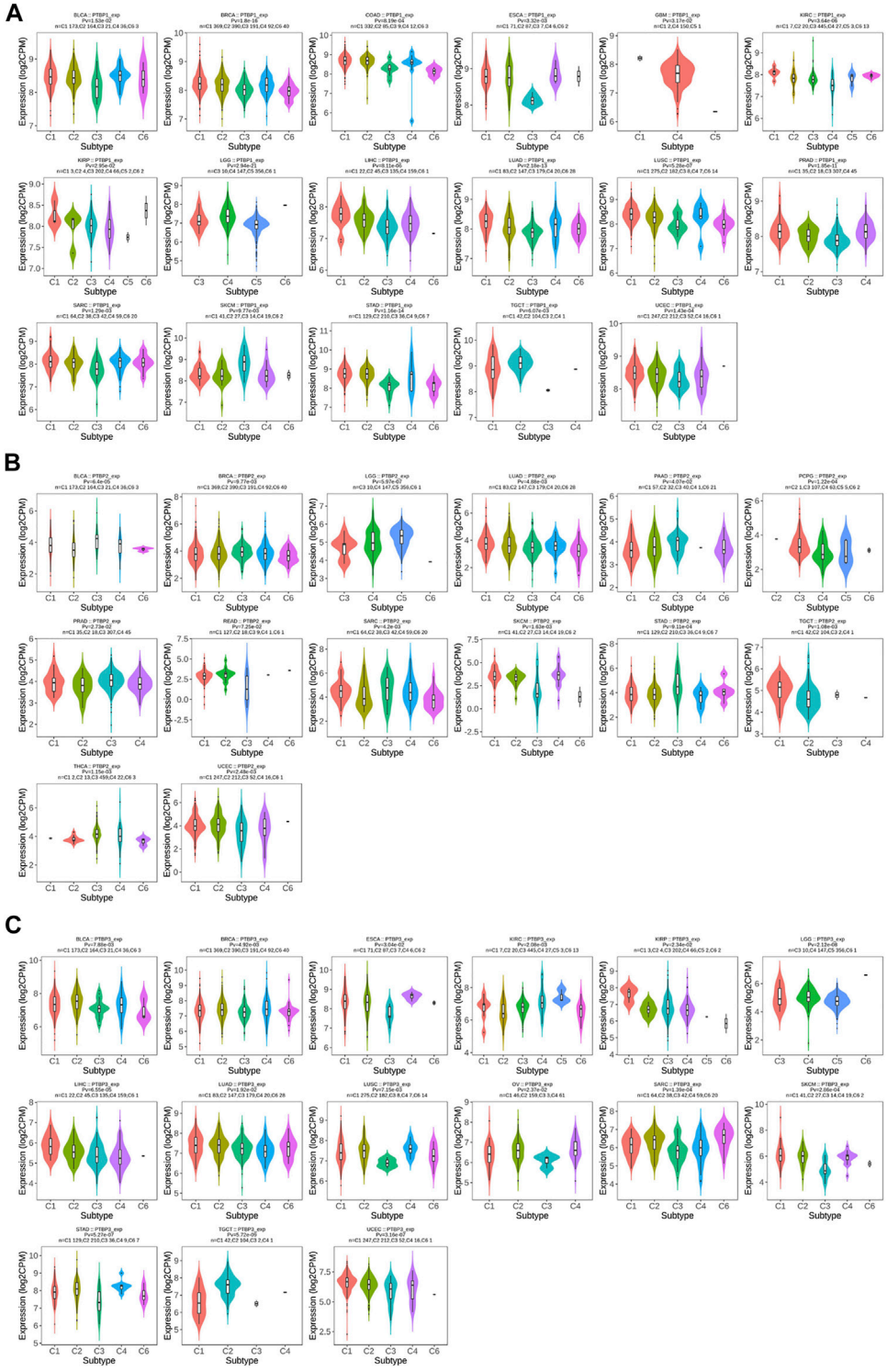


FIGURE 5
The relationship between *PTBP* expression and tumor immune subtype. **(A)** Expression levels of *PTBP1* in immune subtypes. **(B)** Expression levels of *PTBP2* in immune subtypes. **(C)** Expression levels of *PTBP3* in immune subtypes. These results are statistically significant.

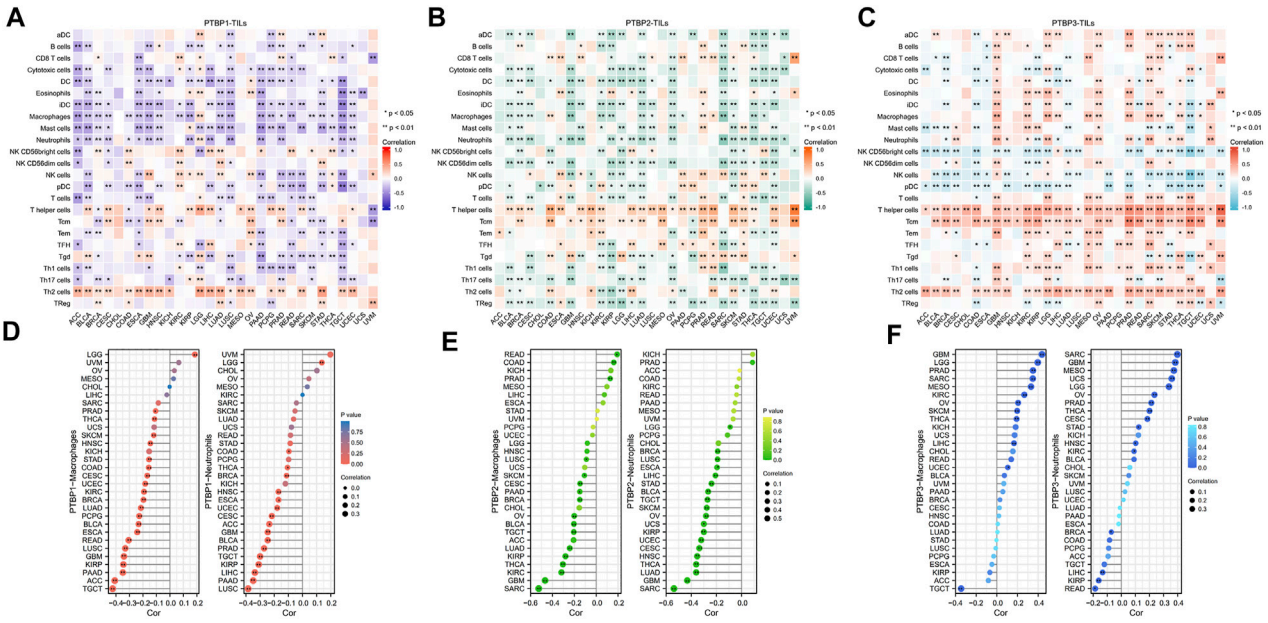


FIGURE 6
Correlation of *PTBP* expression with tumor-infiltrating lymphocytes in the tumor microenvironment in pan-cancer. **(A–C)** Heatmap of the correlation of *PTBP* expression with 24 TILs in pan-cancer. **(D–F)** The correlation of *PTBP* expression with macrophages and neutrophils in various tumors is shown in detail in lollipop plots (*$p < 0.05$; **$p < 0.01$).

explore the potential of *PTBPs* in immunotherapy, we analyzed the relationship between *PTBPs* and four ICP genes in pan-cancer (Supplementary Table S5). In the 30 tumor types, *PTBP1* expression had the most prevalent positive correlation with *LAG3* and *PD-1* expression (in Figure 7A). Similarly, *PTBP2* was generally positively correlated with *CTLA4* and PTBP3 with *CTLA4* and *TIM-3*. These results suggest that *PTBPs* may serve as potential targets for immunotherapy.

PTBP expression correlates with tumor mutational burden and microsatellite instability and can predict immunotherapy efficacy

TMB and MSI are demonstrated biomarkers that predict the efficacy of immunotherapy, with higher TMB or MSI indicating a better response to ICP inhibitors (Chan et al., 2019; Diao et al., 2021). Using the criteria of $|R| \geq 0.3$ and $p < 0.05$, the radar chart showed that the expression of *PTBP1* in ACC, LGG, MESO, and STAD was positively correlated with TMB. *PTBP2* expression was negatively correlated with TMB in UVM. *PTBP3* expression was positively correlated with TMB in ACC and STAD (Figure 7B).

In KICH and LUSC, *PTBP1* expression associated favorably with MSI, but it correlated negatively in READ. *PTBP2* expression was positively correlated with MSI in READ (Figure 7C). The detailed expression data were presented in Supplementary Table S6, and the correlations were shown in Supplementary Table S7. These evidences supported the finding that *PTBPs* may predict response to immunotherapy and play a role in tumor immunity.

Correlation analysis of *PTBPs* and chemical drug sensitivity

We next analyzed the Pearson correlation of *PTBP* expression with the sensitivity of 263 FDA-approved drugs in 60 tumor cell lines using the Cellminer database (Figures 8A–C) and obtained the top six drugs with the strongest correlation with *PTBPs*. For example, the expression of *PTBP1* was proportional to the sensitivity of cells to gemcitabine ($R = 0.409$, $p = 0.001$): the higher the expression of *PTBP1*, the more sensitive the cell was to gemcitabine. Therefore, the expression of *PTBPs* may be a predictor of tumor response to chemotherapeutic drugs.

The analysis of protein-protein interaction

We mapped the PPI networks of PTBP1, PTBP2, and PTBP3 (Figure 9A) respectively and visualized the interaction among these three molecules using STRING (Figure 9B). It showed that PTBP1 was closely related to heterogeneous nuclear ribonucleoproteins (hnRNPs), YBX1, and SFPQ (Meissner et al.,

2000; King et al., 2014). There are four relationships between PTBP1 and PTBP2: experimentally determined interactions, databases recorded interactions, protein homology, and text mining. However, PTBP3 did not seem to interact with PTBP1 and PTBP2 (Figure 9B).

Functional enrichment analysis of *PTBP*-related genes

Spearman correlation analysis on the pan-cancer dataset from cBioPortal yielded 926 mRNAs co-expressed with PTBP1, 657 mRNAs co-expressed with PTBP2, and 874 mRNAs co-expressed with PTBP3 ($|R| \geq 0.4$, $p < 0.05$) (Supplementary Table S8). We further analyzed *PTBP*-related mRNAs using GO (including BP, CC, and MF) and KEGG enrichment analyses (Figures 9C–E). The results revealed that PTBP1 may function through "Cell cycle," "Human T-cell leukemia virus one infection," "RNA transport," "Spliceosome," "DNA replication," "Cellular senescence" and "Apoptosis." PTBP2 may be associated with "Herpes simplex virus one infection" and "Spliceosome" pathways, while PTBP3 may affect tumor progression through "Amyotrophic lateral sclerosis," "Viral carcinogenesis," "Cell cycle" and "Homologous recombination" pathways

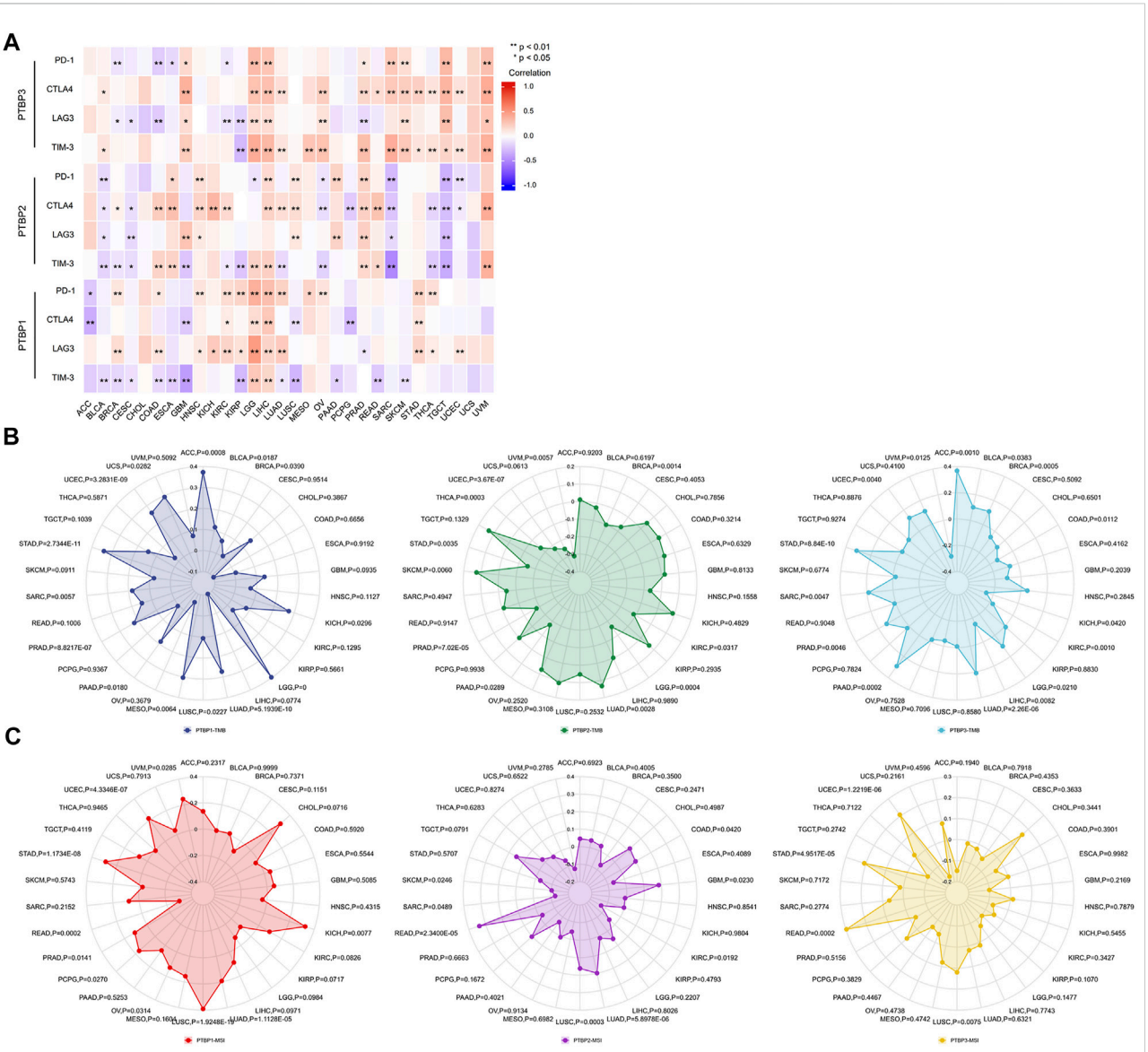


FIGURE 7
Spearman correlation analysis of *PTBP* expression with immune checkpoint genes, tumor mutational burden and microsatellite instability in pan-cancer. **(A)** Correlation of *PTBP* expression with ICPs (PD-1, CTLA4, LAG3 and TIM-3) in pan-cancer (*$p < 0.05$; **$p < 0.01$). **(B)** Correlation of *PTBP* expression with TMB. **(C)** Correlation of *PTBP* expression with MSI. *p* values are marked in the figures.

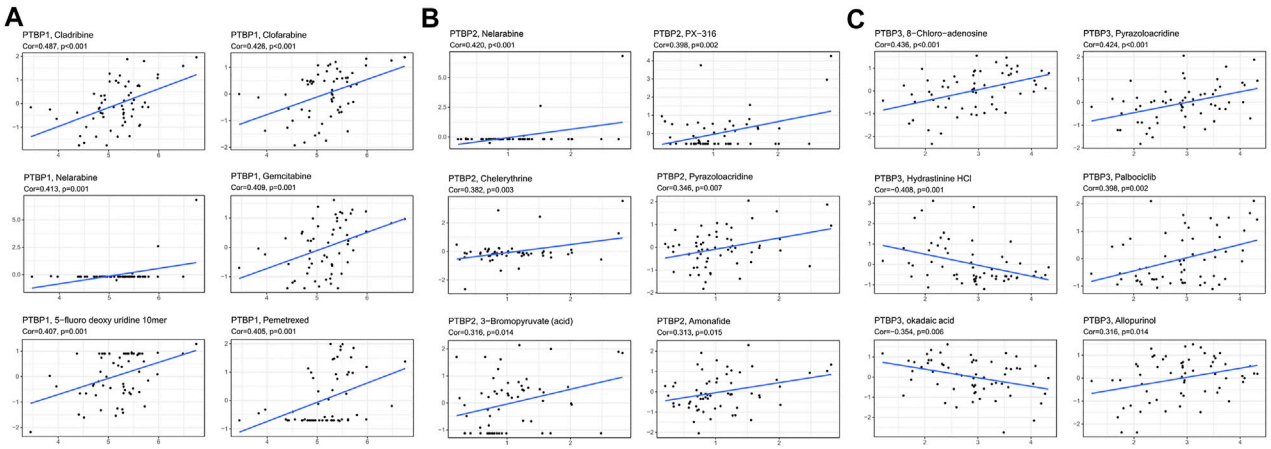


FIGURE 8
Pearson correlation of *PTBP* expression with drug sensitivity scores in various tumor cell lines in Cellminer. The top six drugs with the largest absolute value of the correlation coefficient are displayed. **(A)** Correlation of *PTBP1* with drug sensitivity. **(B)** Correlation of *PTBP2* with drug sensitivity. **(C)** Correlation of *PTBP3* with drug sensitivity. The correlation coefficient and *p* value are marked in the figure.

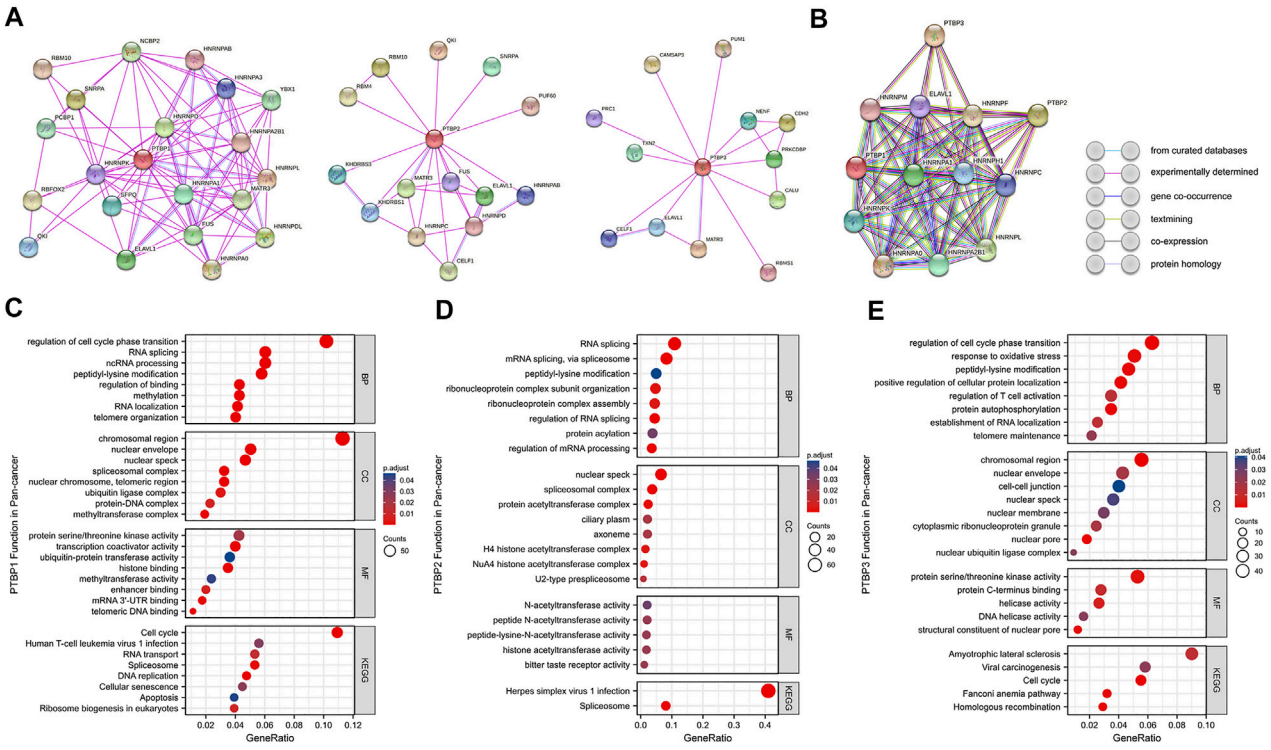


FIGURE 9
Protein-protein interaction networks and functional enrichment analysis of PTBPs in pan-cancer. **(A)** Experimentally validated interacting proteins of PTBPs using STRING. **(B)** The interaction relationship among the three protein molecules. Line colors in the legend indicate different relationships. **(C–E)** GO enrichment and KEGG pathway analysis results for co-expressed mRNAs of *PTBPs*.

(Supplementary Table S9). Together, the results above have established a novel theoretical framework for the investigation of PTBP regulation mechanism in malignancies.

Discussion

PTBPs are RNA-binding proteins that are involved in alternative splicing, mRNA stability, and translation. The PTBP family includes PTBP1, PTBP2 and PTBP3. PTBP1 can be expressed in almost all types of cells; PTBP2 is only expressed in the neurological system while PTBP3 is found mostly in immune cells (Spellman et al., 2007). Among the PTBP family members, PTBP1 is most frequently linked with cancer, followed by PTBP3. It is reported that PTBP1 promotes lung cancer metastasis by regulating the alternative splicing of *Mena* mRNA (Li et al., 2019). PTBP3 is upregulated in breast cancer and regulates *ZEB1* mRNA stability to promote epithelial-mesenchymal transition in BRCA (Hou et al., 2018; Liang et al., 2020). Most research has focused on the function of PTBPs in tumor cells, but little attention has been paid to their interaction with immune cells in the TME (Sasanuma et al., 2019; Geng et al., 2021). In addition, reports of PTBPs in uncommon tumors are rare. Thus, we investigated the expression, function, and immune characterization of PTBP1, PTBP2, and PTBP3 in pan-cancer.

We first performed differentiation analysis and correlation analysis on the expression of the *PTBP1*, *PTBP2*, and *PTBP3* genes in 30 tumor types using TCGA. The results showed that in most tumor types, the expression levels of *PTBP1* and *PTBP3* in tumor tissues were significantly higher than that in non-tumor tissues. In contrast, the expression level of *PTBP2* was lower in tumor tissues compared with that in normal tissues. Interestingly, when we analyzed the expression correlation of *PTBPs*, we found that *PTBP1/3* appeared to have opposite expression trends to *PTBP2* in pan-cancer. Our results were consistent with other scholars' findings. Previous studies suggested that PTBP1 is a repressor of PTBP2 and that there was a "switch" between the two molecules (Boutz et al., 2007; Spellman et al., 2007). SON may be an on-off regulator of the expression of *PTBP1* and *PTBP2* in GBM (Kim et al., 2021), and PTBP2 compensates for the absence of *Ptbp1* during B cell development in mice (Monzón-Casanova et al., 2020). Furthermore, we found a co-expression trend between *PTBP1* and *PTBP3* which deserved to be further investigated.

It has been reported in the literature that PTBP1 can be used as a biomarker for poor prognosis in bladder cancer (Bielli et al., 2018), and PTBP3 as a therapeutic target for gastric cancer (Liang et al., 2017). Here we further comprehensively explored the association of *PTBP* expression with prognosis in pan-cancer. Through multivariate Cox regression analysis and OS-KM survival curves, we found that patients with ACC, LGG, and PAAD had poor prognosis when *PTBP3* was highly expressed, but patients with KIRC had better prognosis. Given that the expression of *PTBP3* in KIRC tumor tissues was significantly lower than that in control tissues, PTBP3 may be a tumor suppressor molecule in KIRC. Thus, more research is required to examine the function and molecular mechanism of PTBP3 in KIRC. Time-dependent ROC curves were used to verify the prognostic value of *PTBPs* in pan-cancer. Compared with the ordinary ROC curve, the time-dependent ROC curve detects the accuracy of candidate markers at specified times. We finally identified *PTBP1* in ACC, KIRP, and LGG; *PTBP2* in ACC and KICH; and *PTBP3* in ACC, LGG, and PAAD as potential prognostic biomarkers that may be involved in tumor progression in these tumor types.

We then analyzed the expression of *PTBPs* in different immune subtypes. The results indicated that *PTBPs* might participate in immune regulation. The expression of *PTBPs* was significantly different across multiple immune subtypes and strongly correlated with the number of TILs in the TME. Remarkably, *PTBP1* on Th2 cells, *PTBP2* on T helper cells and Tcm, and *PTBP3* on T helper cells, Tcm, and Th2 cells may have broad positive regulatory effects in pan-cancer. *PTBPs* are also strongly associated with macrophages and neutrophils in the TME. For example, *PTBP3* expression was positively correlated with macrophages and neutrophils in GBM, LGG, PRAD, SARC, MESO, KIRC, OV, and THCA. These results demonstrated the important role of PTBPs in tumor immunity and the tumor microenvironment.

The immune microenvironment in tumor tissues leads to tumor heterogeneity, which influences the clinical efficacy of anticancer drugs. Immune checkpoint inhibitors are used as treatment options for cancer patients. We found that *PTBP* expression showed a strong correlation with *PD-1*, *CTLA4*, *LAG3*, or *TIM-3* in pan-cancer. Therefore, *PTBPs* may be a class of potential therapeutic targets, providing a new direction for combined targeted immunotherapy in the future.

We also analyzed the correlation of *PTBPs* with TMB and MSI. Tumor cells with high TMB usually have higher levels of neoantigens, which help the immune system to recognize the tumor and activate the anti-tumor effect of T cells. Therefore, higher TMB generally indicates better outcome of immunotherapy, and TMB is highly correlated with the efficacy of PD-1/PD-L1 inhibitors (Yarchoan et al., 2017; Chan et al., 2019). MSI works similarly. TMB and MSI have become predictive markers of tumor immunotherapy efficacy in recent years. The correlation of *PTBP* expression with TMB and MSI in pan-cancer suggests that PTBPs may become novel biomarkers for predicting patients' response to immunotherapy.

We also made other notable findings. Gemcitabine is an effective anti-tumor drug for NSCLC (stage III and IV), OV, BRCA, BLCA, and other malignant tumors (Ferrazzi and Stievano, 2006; Mornex and Girard, 2006), and 5-fluorodeoxyuridine is a common chemotherapeutic drug for BRCA, STAD, READ, and BLCA (Koizumi et al., 1993). *PTBP1* expression was proportional to the sensitivity of cells to gemcitabine ($R = 0.409$, $p = 0.001$) and 5-fluorodeoxyuridine ($R = 0.407$, $p = 0.001$). This result indicates that

the expression of *PTBPs* may predict the therapeutic effect of chemotherapeutic drugs.

When we investigated "interacting proteins", we found that all PTBP family proteins can tightly interacts with hnRNPs and ELAVL1 (also known as HuR), which was verified by co-immunoprecipitation or reported in literature (Hegele et al., 2012). The presence of such protein complexes may increase their effect. For example, PTBP1 can interact with HuR and jointly upregulate the translation of *HIF-1α* mRNA in human cervical carcinoma HeLa cells (Galbán et al., 2008).

In the enrichment analysis of co-expressed genes, we inferred that PTBPs may function in the cell cycle, RNA splicing and RNA localization. PTBP1 and PTBP3 were enriched in telomere-related signaling pathways, suggesting that they may be involved in cellular senescence pathways. Scientists found that PTBP1 can regulate alternative splicing of genes involved in intracellular trafficking to control the senescence-associated secretory phenotype (SASP). Inhibition of PTBP1 blocks the tumor-promoting effect of SASP and impair immune surveillance (Georgilis et al., 2018). Sayed et al. also found that knockdown of *PTBP1* in cancer cells reduced *hTERT* full-length splicing and telomerase activity (Sayed et al., 2019). The important role of PTBP1 and PTBP3 in cellular senescence and immunity should be further explored.

This study has several limitations. First, our conclusions are limited by sequencing technologies and analytical methodologies from the database, and the data may be lacking in granularity and precision. This has become a pervasive problem in bioinformatics research. Second, whether *PTBPs* can be used as biomarkers for prognosis and immunotherapy requires validation in more clinical samples. At present, there is no immune-targeted drug against PTBPs, so it is not possible to clinically verify the effect of these targets. Third, the involvement of PTBPs in immune regulation and cellular senescence need to be supported by *in vitro* and *in vivo* experimental evidence.

Conclusion

This study comprehensively and systematically analyzed the prognostic value, genetic variation, and signaling pathways of PTBP1, PTBP2, and PTBP3 and the correlation of *PTBP* expression with TILs, ICP, TMB, MSI, and drug sensitivity from a pan-cancer perspective. Our results indicate that *PTBPs* may be promising prognostic biomarkers and predict the response to immunotherapy in pan-cancer. We found that PTBPs are closely related to tumor progression and cell senescence, which provides a theoretical reference for subsequent research.

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

CC, AS, and DL designed the research study. CC, YG, JH, and GL carried out the data analysis and processing. CC prepared the original manuscript. AS, WC, and DL revised the manuscript and edited the final version. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (82072362, 8200223, 8200222); the Postdoctoral Science Foundation of China (2020M681399); the Science and Technology Innovation Action Plan of Shanghai Science and Technology Commission (19411964800, 20YF1443700); Shanghai Municipal Health and Health Commission (20204Y0070); Shanghai Public Health System Construction Three-Year Action Plan (2020–2022) Key Disciplines (GWV-10.1-XK04) and the Clinical Research Project of Tongji Hospital of Tongji University [ITJ (ZD) 1905, ITJ (QN)1905, ITJ (QN)2005].

Acknowledgments

We would like to acknowledge all members of the "HOME for Researchers" for their assistance and useful suggestions.

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 FY declared a shared parent 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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.968458/full#supplementary-material

References

Bielli, P., Panzeri, V., Lattanzio, R., Mutascio, S., Pieraccioli, M., Volpe, E., et al. (2018). The splicing factor PTBP1 promotes expression of oncogenic splice variants and predicts poor prognosis in patients with non-muscle-invasive bladder cancer. *Clin. Cancer Res.* 24, 5422–5432. doi:10.1158/1078-0432.CCR-17-3850

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Bonneville, R., Krook, M. A., Kautto, E. A., Miya, J., Wing, M. R., Chen, H. Z., et al. (2017). Landscape of microsatellite instability across 39 cancer types. *JCO Precis. Oncol.* 1, 1–15. doi:10.1200/PO.17.00073

Boutz, P. L., Stoilov, P., Li, Q., Lin, C. H., Chawla, G., Ostrow, K., et al. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* 21, 1636–1652. doi:10.1101/gad.1558107

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol.* 30, 44–56. doi:10.1093/annonc/mdy495

Diao, Z., Han, Y., Chen, Y., Zhang, R., and Li, J. (2021). The clinical utility of microsatellite instability in colorectal cancer. *Crit. Rev. Oncol. Hematol.* 157, 103171. doi:10.1016/j.critrevonc.2020.103171

Donlon, N. E., Power, R., Hayes, C., Reynolds, J. V., and Lysaght, J. (2021). Radiotherapy, immunotherapy, and the tumour microenvironment: Turning an immunosuppressive milieu into a therapeutic opportunity. *Cancer Lett.* 502, 84–96. doi:10.1016/j.canlet.2020.12.045

Ferrazzi, E., and Stievano, L. (2006). Gemcitabine: monochemotherapy of breast cancer. *Ann. Oncol.* 17, v169–172. doi:10.1093/annonc/mdj975

Gaikwad, S., Agrawal, M. Y., Kaushik, I., Ramachandran, S., and Srivastava, S. K. (2022). Immune checkpoint proteins: Signaling mechanisms and molecular interactions in cancer immunotherapy. *Semin. Cancer Biol.* doi:10.1016/j.semcancer.2022.03.014

Galbán, S., Kuwano, Y., Pullmann, R., Jr, Martindale, J. L., Kim, H. H., Lal, A., et al. (2008). RNA-binding proteins HuR and PTB promote the translation of hypoxia-inducible factor 1alpha. *Mol. Cell. Biol.* 28, 93–107. doi:10.1128/MCB.00973-07

Geng, G., Xu, C., Peng, N., Li, Y., Liu, J., Wu, J., et al. (2021). PTBP1 is necessary for dendritic cells to regulate T-cell homeostasis and antitumour immunity. *Immunology* 163, 74–85. doi:10.1111/imm.13304

Georgilis, A., Klotz, S., Hanley, C. J., Herranz, N., Weirich, B., Morancho, B., et al. (2018). PTBP1-Mediated alternative splicing regulates the inflammatory secretome and the pro-tumorigenic effects of senescent cells. *Cancer Cell* 34, 85–102.e9. doi:10.1016/j.ccell.2018.06.007

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., et al. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell* 45, 567–580. doi:10.1016/j.molcel.2011.12.034

Hou, P., Li, L., Chen, F., Chen, Y., Liu, H., Li, J., et al. (2018). PTBP3-Mediated regulation of ZEB1 mRNA stability promotes epithelial-mesenchymal transition in breast cancer. *Cancer Res.* 78, 387–398. doi:10.1158/0008-5472.CAN-17-0883

Ji, Q., Zhang, L., Liu, X., Zhou, L., Wang, W., Han, Z., et al. (2014). Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. *Br. J. Cancer* 111, 736–748. doi:10.1038/bjc.2014.383

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Keenan, T. E., Burke, K. P., and Van Allen, E. M. (2019). Genomic correlates of response to immune checkpoint blockade. *Nat. Med.* 25, 389–402. doi:10.1038/s41591-019-0382-x

Kim, J. H., Jeong, K., Li, J., Murphy, J. M., Vukadin, L., Stone, J. K., et al. (2021). SON drives oncogenic RNA splicing in glioblastoma by regulating PTBP1/PTBP2 switching and RBFOX2 activity. *Nat. Commun.* 12, 5551. doi:10.1038/s41467-021-25892-x

King, H. A., Cobbold, L. C., Pichon, X., Pöyry, T., Wilson, L. A., Booden, H., et al. (2014). Remodelling of a polypyrimidine tract-binding protein complex during apoptosis activates cellular IRESs. *Cell Death Differ.* 21, 161–171. doi:10.1038/cdd.2013.135

Koizumi, W., Kurihara, M., Sasai, T., Yoshida, S., Morise, K., Imamura, A., et al. (1993). A phase II study of combination therapy with 5'-deoxy-5-fluorouridine and cisplatin in the treatment of advanced gastric cancer with primary foci. *Cancer* 72, 658–662. doi:10.1002/1097-0142(19930801)72:3<658::aid-cncr2820720306>3.0.co;2-k

König, I. R., Fuchs, O., Hansen, G., von Mutius, E., and Kopp, M. V. (2017). What is precision medicine. *Eur. Respir. J.* 50, 1700391. doi:10.1183/13993003.00391-2017

Li, S., Shen, L., Huang, L., Lei, S., Cai, X., Breitzig, M., et al. (2019). PTBP1 enhances exon11a skipping in Mena pre-mRNA to promote migration and invasion in lung carcinoma cells. *Biochim. Biophys. Acta. Gene Regul. Mech.* 1862, 858–869. doi:10.1016/j.bbagrm.2019.04.006

Liang, X., Shi, H., Yang, L., Qiu, C., Lin, S., Qi, Y., et al. (2017). Inhibition of polypyrimidine tract-binding protein 3 induces apoptosis and cell cycle arrest, and enhances the cytotoxicity of 5- fluorouracil in gastric cancer cells. *Br. J. Cancer* 116, 903–911. doi:10.1038/bjc.2017.32

Liang, Y., Song, X., Li, Y., Chen, B., Zhao, W., Wang, L., et al. (2020). LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis. *Mol. Cancer* 19, 85. doi:10.1186/s12943-020-01206-5

Meissner, M., Dechat, T., Gerner, C., Grimm, R., Foisner, R., and Sauermann, G. (2000). Differential nuclear localization and nuclear matrix association of the splicing factors PSF and PTB. *J. Cell. Biochem.* 76, 559–566. doi:10.1002/(sici)1097-4644(20000315)76:4<559::aid-jcb4>3.0.co;2-u

Mickleburgh, I., Kafasla, P., Cherny, D., Llorian, M., Curry, S., Jackson, R. J., et al. (2014). The organization of RNA contacts by PTB for regulation of FAS splicing. *Nucleic Acids Res.* 42, 8605–8620. doi:10.1093/nar/gku519

Monzón-Casanova, E., Matheson, L. S., Tabbada, K., Zarnack, K., Smith, C. W., and Turner, M. (2020). Polypyrimidine tract-binding proteins are essential for B cell development. *Elife* 9, e53557. doi:10.7554/eLife.53557

Mornex, F., and Girard, N. (2006). Gemcitabine and radiation therapy in non-small cell lung cancer: state of the art. *Ann. Oncol.* 17, 1743–1747. doi:10.1093/annonc/mdl117

Newnes, H. V., Armitage, J. D., Audsley, K. M., Bosco, A., and Waithman, J. (2021). Directing the future breakthroughs in immunotherapy: The importance of a holistic approach to the tumour microenvironment. *Cancers (Basel)* 13, 5911. doi:10.3390/cancers13235911

Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., et al. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* 309, 2054–2057. doi:10.1126/science.1114066

Ribas, A., and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science* 359, 1350–1355. doi:10.1126/science.aar4060

Sasanuma, H., Ozawa, M., and Yoshida, N. (2019). RNA-binding protein Ptbp1 is essential for BCR-mediated antibody production. *Int. Immunol.* 31, 157–166. doi:10.1093/intimm/dxy077

Sayed, M. E., Yuan, L., Robin, J. D., Tedone, E., Batten, K., Dahlson, N., et al. (2019). NOVA1 directs PTBP1 to hTERT pre-mRNA and promotes telomerase activity in cancer cells. *Oncogene* 38, 2937–2952. doi:10.1038/s41388-018-0639-8

Singh, R., Valcárcel, J., and Green, M. R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* 268, 1173–1176. doi:10.1126/science.7761834

Spellman, R., Llorian, M., and Smith, C. W. (2007). Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* 27, 420–434. doi:10.1016/j.molcel.2007.06.016

Sun, S., Xu, L., Zhang, X., Pang, L., Long, Z., Deng, C., et al. (2021). Systematic assessment of transcriptomic biomarkers for immune checkpoint blockade response in cancer immunotherapy. *Cancers (Basel)* 13, 1639. doi:10.3390/cancers13071639

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Tan, L. Y., Whitfield, P., Llorian, M., Monzon-Casanova, E., Diaz-Munoz, M. D., Turner, M., et al. (2015). Generation of functionally distinct isoforms of PTBP3 by alternative splicing and translation initiation. *Nucleic Acids Res.* 43, 5586–5600. doi:10.1093/nar/gkv429

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The immune landscape of cancer. *Immunity* 48, 812–830.e14. doi:10.1016/j.immuni.2018.03.023

Tian, J., Liu, Y., Zhang, T., Yue, L., Xiao, Y., and Guo, C. (2022). LAG-3 is a promising inhibitory immune checkpoint for antitumor immunotherapy. *Expert Rev. Anticancer Ther.* 22, 289–296. doi:10.1080/14737140.2022.2039124

Tu, L., Guan, R., Yang, H., Zhou, Y., Hong, W., Ma, L., et al. (2020). Assessment of the expression of the immune checkpoint molecules PD-1, CTLA4, TIM-3 and LAG-3 across different cancers in relation to treatment response, tumor-infiltrating immune cells and survival. *Int. J. Cancer* 147, 423–439. doi:10.1002/ijc.32785

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi:10.1093/nar/gkab1061

Yang, J., Zhao, S., Wang, J., Sheng, Q., Liu, Q., and Shyr, Y. (2021). A pan-cancer immunogenomic atlas for immune checkpoint blockade immunotherapy. *Cancer Res.* 82, 539–542. doi:10.1158/0008-5472.CAN-21-2335

Yarchoan, M., Hopkins, A., and Jaffee, E. M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* 377, 2500–2501. doi:10.1056/NEJMc1713444

frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 29 August 2022
DOI 10.3389/fphys.2022.918381

Check for updates

OPEN ACCESS

EDITED BY
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

REVIEWED BY
Guang Yang,
Imperial College London, United Kingdom
Sandra Ortega-Martorell,
Liverpool John Moores University, United Kingdom

*CORRESPONDENCE
Yongjie Wang,
yongjiewang@csu.edu.cn
Longxiang Wu,
wulongxiang1123@126.com

SPECIALTY SECTION
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

RECEIVED 12 April 2022
ACCEPTED 28 July 2022
PUBLISHED 29 August 2022

CITATION
Yi Z, Ou Z, Hu J, Qiu D, Quan C, Othmane B, Wang Y and Wu L (2022), Computer-aided diagnosis of prostate cancer based on deep neural networks from multi-parametric magnetic resonance imaging.
Front. Physiol. 13:918381.
doi: 10.3389/fphys.2022.918381

COPYRIGHT
© 2022 Yi, Ou, Hu, Qiu, Quan, Othmane, Wang 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.

Computer-aided diagnosis of prostate cancer based on deep neural networks from multi-parametric magnetic resonance imaging

Zhenglin Yi[1,2], Zhenyu Ou[1,2], Jiao Hu[1,2], Dongxu Qiu[1,2], Chao Quan[1,2], Belaydi Othmane[1], Yongjie Wang[3,2]* and Longxiang Wu[1,2]*

[1]Department of Urology, Xiangya Hospital, Central South University, Changsha, China, [2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, [3]Department of Burns and Plastic Surgery, Xiangya Hospital, Central South University, Changsha, China

Objectives: To evaluate a new deep neural network (DNN)–based computer-aided diagnosis (CAD) method, namely, a prostate cancer localization network and an integrated multi-modal classification network, to automatically localize prostate cancer on multi-parametric magnetic resonance imaging (mp-MRI) and classify prostate cancer and non-cancerous tissues.

Materials and methods: The PROSTAREx database consists of a "training set" (330 suspected lesions from 204 cases) and a "test set" (208 suspected lesions from 104 cases). Sequences include T2-weighted, diffusion-weighted, Ktrans, and apparent diffusion coefficient (ADC) images. For the task of abnormal localization, inspired by V-net, we designed a prostate cancer localization network with mp-MRI data as input to achieve automatic localization of prostate cancer. Combining the concepts of multi-modal learning and ensemble learning, the integrated multi-modal classification network is based on the combination of mp-MRI data as input to distinguish prostate cancer from non-cancerous tissues through a series of operations such as convolution and pooling. The performance of each network in predicting prostate cancer was examined using the receiver operating curve (ROC), and the area under the ROC curve (AUC), sensitivity (TPR), specificity (TNR), accuracy, and Dice similarity coefficient (DSC) were calculated.

Results: The prostate cancer localization network exhibited excellent performance in localizing prostate cancer, with an average error of only 1.64 mm compared to the labeled results, an error of about 6%. On the test dataset, the network had a sensitivity of 0.92, specificity of 0.90, PPV of 0.91, NPV of 0.93, and DSC of 0.84. Compared with multi-modal classification networks, the performance of single-modal classification networks is slightly inadequate. The integrated multi-modal classification network performed best in classifying prostate cancer and non-cancerous tissues with a TPR of 0.95, TNR of 0.82, F1-Score of 0.8920, AUC of 0.912, and accuracy of 0.885, which fully confirmed the feasibility of the ensemble learning approach.

Conclusion: The proposed DNN-based prostate cancer localization network and integrated multi-modal classification network yielded high performance in experiments, demonstrating that the prostate cancer localization network and integrated multi-modal classification network can be used for computer-aided diagnosis (CAD) of prostate cancer localization and classification.

KEYWORDS

deep neural networks (DNN), computer-aided diagnosis (CAD), prostate cancer localization, prostate cancer classification, multi-parametric magnetic resonance imaging (MP-MRI)

1 Introduction

Prostate cancer is the most common malignant tumor of the male genitourinary system and has become the second most common malignant tumor in men worldwide, second only to lung cancer (Sung et al., 2021). Image information is of great significance for the diagnosis of prostate cancer. Transrectal prostate color Doppler ultrasound can be used as a screening tool for prostate cancer. Magnetic resonance examination is widely used to evaluate prostate cancer, and pathological examination and Gleason score are an important basis for prostate grading (Litwin and Tan, 2017). In the clinical diagnosis of prostate cancer, a radiologist is required to separate the prostate tissue from the surrounding tissues and organs in the prostate MRI. The meaningful information extracted by this segmentation process includes shape, the relative position of organs, volume, and abnormal signals. Because the area of prostate tissue in MRI is small, less valid information is available, and the size, shape, and location of prostate tissue vary from patient to patient. Precise localization of the prostate and identification of prostate cancer remains difficult for radiologists.

In recent years, deep learning technology has developed rapidly in the medical field, which can extract features from image data in a supervised or unsupervised manner for image classification or segmentation. Deep neural network (DNN) is an artificial neural network that imitates the function of human neurons and can perform tasks such as classification (Ciresan et al., 2012), image segmentation (Quan et al., 2021), and entity reconstruction (Nguyen et al., 2020). It has a stronger expressive ability and can fit almost any function, but it also has problems such as many network parameters, a large training amount, and difficulty in training. The specific structure of DNN is shown in Supplementary Figure S1. The use of DNNs is growing exponentially, and researchers have used DNNs to correctly classify a large number of different classes of images (Deng et al., 2009). One of the main uses of DNN in medicine is to aid in the diagnosis of certain types of cancer, which are often identified clinically by skilled radiologists from medical images. Cancer detection methods based on artificial intelligence and MRI are widely used in daily clinical diagnosis, which has achieved higher diagnostic success rates than experienced radiologists. The study shows that the success rate of lung cancer detection and breast cancer detection using DNN is significantly better than the results of manual detection by radiologists (Becker et al., 2017; Coudray et al., 2018). In addition to the use of DNN on MRI, other studies have shown that the use of DNN can help determine the accuracy of results from transrectal biopsies for the diagnosis of prostate cancer (Takeuchi et al., 2019).

Therefore, DNN-based automatic localization of prostate MRI and prostate cancer diagnosis is a study with good clinical application prospects, which can assist radiologists in better diagnosis of prostate cancer. We propose a computer-aided diagnosis (CAD) method for localization and classification of prostate cancer based on DNN and mp-MRI, called prostate cancer localization network and integrated multi-modal classification network, aiming to improve the efficiency of radiologists. First, we used MR images provided by public prostate cancer databases and preprocessed them for training localization and classification models (Armato et al., 2018). The previously defined metric algorithm was then fully evaluated using a test set and a validation set. Finally, the output contains the image locations of possible malignancies and the likelihood of detecting prostate cancer based on the patient's multi-parametric MRI.

2 Materials and methods

The goal of this study was to propose a new diagnostic assistant technology for prostate cancer, which uses multiple prostate MRIs of each patient as the input of the localization and classification model, and the output is the specific location of the localization and the classification of benign and malignant tissue, aiming to identify potential tumors. It is worth stating that our proposed localization network and classification network are both studied independently.

2.1 Data set

The PROSTATEx public database used in this study is part of the SPIE-AAPM-NCI Prostate MR Classification Challenge, which aims to advance the diagnostic classification of prostate cancer by analyzing prostate MRI (Armato et al., 2018). The database, collected by Radboud University Medical Centre (Radboudumc), covers more than 300,000 prostate MRIs from

TABLE 1 Details of PROSTATEx dataset.

Category	PZ	TZ	As	SV	Total
Training set	191	82	55	2	330
Test set	113	59	34	2	208

AS, anterior fibromuscular stroma; PZ, peripheral zone; SV, seminal vesicle; TZ, transitional zone.

346 patients, including T2-weighted image (T2WI), proton density–weighted image (PdWI), dynamic contrast enhancement (DCE), and diffusion-weighted (DW) images. Each patient has a Ktrans image, one or more DW images, and one or more T2 images.

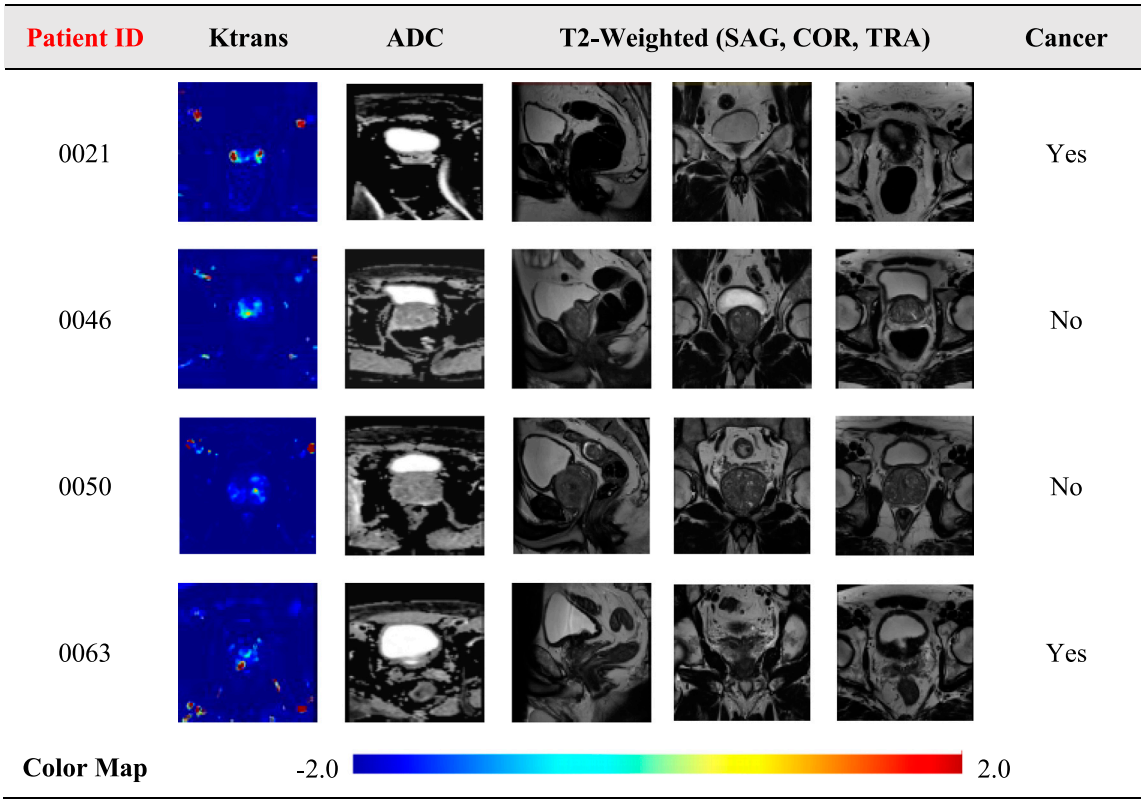
In the DICOM file of the PROSTATEx dataset, the header information consists of the acquisition information of the image with the basic information of the case. Among them, the image acquisition information includes acquisition time, size, repetition time, pixel spacing, image position, and orientation. The dataset provides the coordinates of one or more points of interest (POIs) and information on the prostate area. Based on human experience, the characteristics of prostate lesions vary from region to region. Four prostate zones are associated with the POI provided: the peripheral zone (PZ), the transitional zone (TZ), the anterior fibromuscular stroma (AS), and the seminal vesicle (SV). The details of the dataset are shown in Table 1. The specific image classification of the database is shown in Table 2.

The training set of the PROSTAREx dataset contains 204 cases with 330 suspicious cancer lesions, of which 76 are gold standard "True" lesions, and the remaining 254 are gold standard "False" lesions. The test set contains 104 cases and 208 suspicious lesions to be diagnosed.

2.2 Data preprocessing

We observed the PROSTATEx dataset using ITK-SNAP and concluded that there are three problems in MR image preprocessing of the prostate: 1) abnormal data acquisition, such as missing sequences and different acquisition order; 2) different image resolution and gray value distribution among sequences; 3) insufficient sample size, containing only

TABLE 2 PROSTATEx database image classification, including Ktrans, ADC, and t2-weighted images.



ADC, apparent diffusion coefficient; COR, coronal; SAG, sagittal; TRA, transverse.

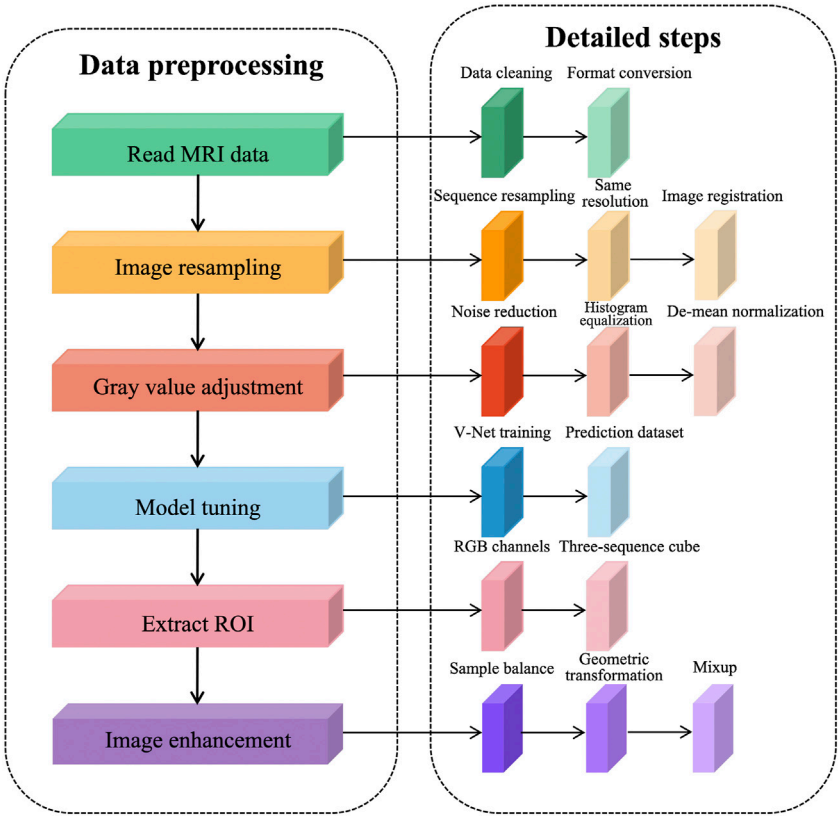


FIGURE 1
Prostate MRI data preprocessing steps.

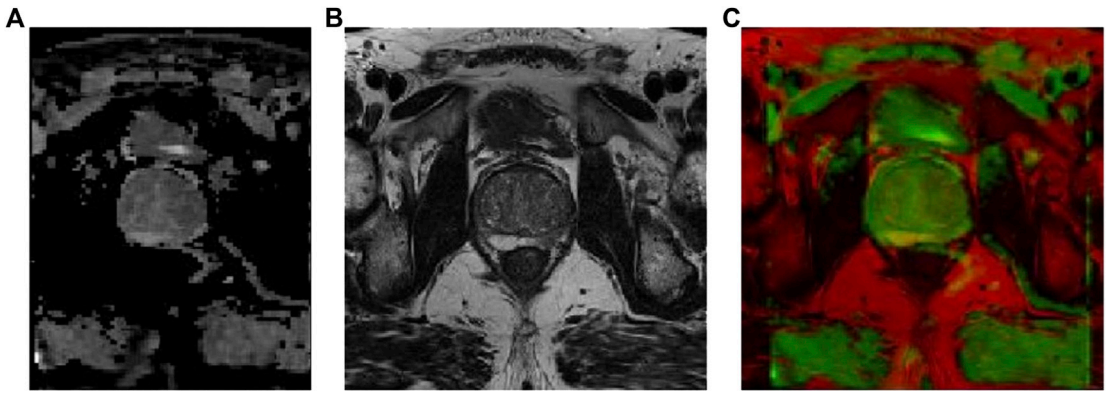


FIGURE 2
Prostate MR image alignment results, where **(A)** ADC image, **(B)** T2-weighted image, and **(C)** overlap map after image alignment.

330 training samples, and the network training is prone to overfitting.

To solve the problems existing in the dataset and further reduce the redundant information of network learning, the following preprocessing steps are proposed in this article, as shown in Figure 1. The prostate MRI data are first read so that the format is consistent between sequences, and then image resampling is performed so that the pixel spacing is consistent

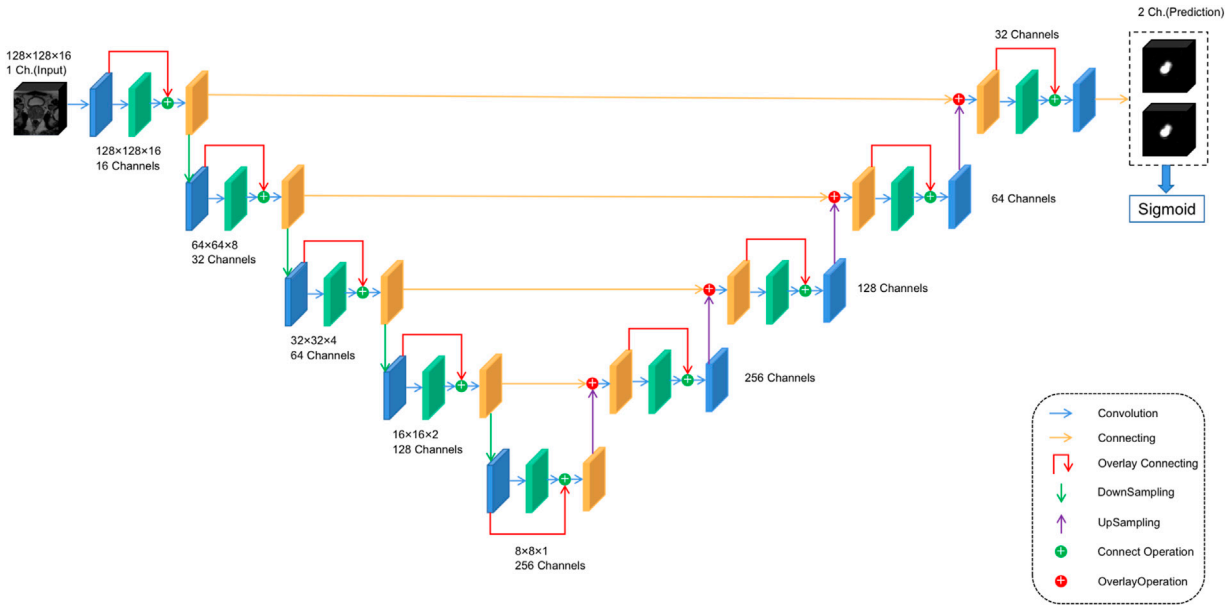


FIGURE 3
Network structure of the V-Net–based prostate cancer anomaly localization system.

from case to case and from sequence to sequence. Next, the grayscale values are adjusted to ensure a consistent histogram distribution for each sequence. To further reduce redundant information, prostate tissue is extracted to reduce the learning of background information by the network. After generating a region of interest (ROI) that matches the network input structure, image enhancement is performed to expand the sample size.

We performed data cleaning on the training set and eliminated the cases with missing sequences. This decision reduced two cases, leaving 202 cases after elimination, each containing four sequences, namely, T2WI sequences, DWI sequences, Ktrans sequences, and ADC sequences. To expand the sample size, we used the common image enhancement methods of flip, pan, rotate, and zoom for multiple images of prostate MRI, and in addition, used the data enhancement method of Mixup to improve the linear expression between different samples.

The prostate alignment transformation used in this study is the B spline transformation. To make the alignment easier, two resolutions are used, first using a low resolution for the alignment and then a high resolution for the alignment. In performing the B spline transformation, a mutual information function with an increased penalty for rigidity is used as the optimization objective using an adaptive gradient descent algorithm. Finally, the rigid transform and B spline transformation were combined to obtain the final transform results. The results of prostate MR image alignment are shown in Figure 2.

2.3 Prostate cancer localization network structure

For anomaly localization, previous research studies have created a 3D convolutional DNN specifically for medical image segmentation. The architecture used in this study is based on V-net (Milletari et al., 2016), a well-known image segmentation network for medical imaging. The main modification made in this study is the redefinition of the input and output tensors and activation functions. Both input and output tensors are of size (128,128,16,1), and Leaky ReLU is used as the activation function because the original PreLU activation function increases the risk of overtraining in small databases. At the same time, the output layer uses the sigmoid activation function, which can get the binary position of the tumor, and the output is reflected as the segmentation of the same position in the figure. The specific architecture is shown in Figure 3 to facilitate its repeatability.

2.4 Single-modal classification network structure

To address abnormal single-modal image classification, a lightweight architecture based on Inception-V3 and VGG-16 networks is proposed (Rueckauer et al., 2017). Typically, in this type of architecture, ReLU is chosen as the activation function (Dahl et al., 2013). However, to avoid problems such as gradient

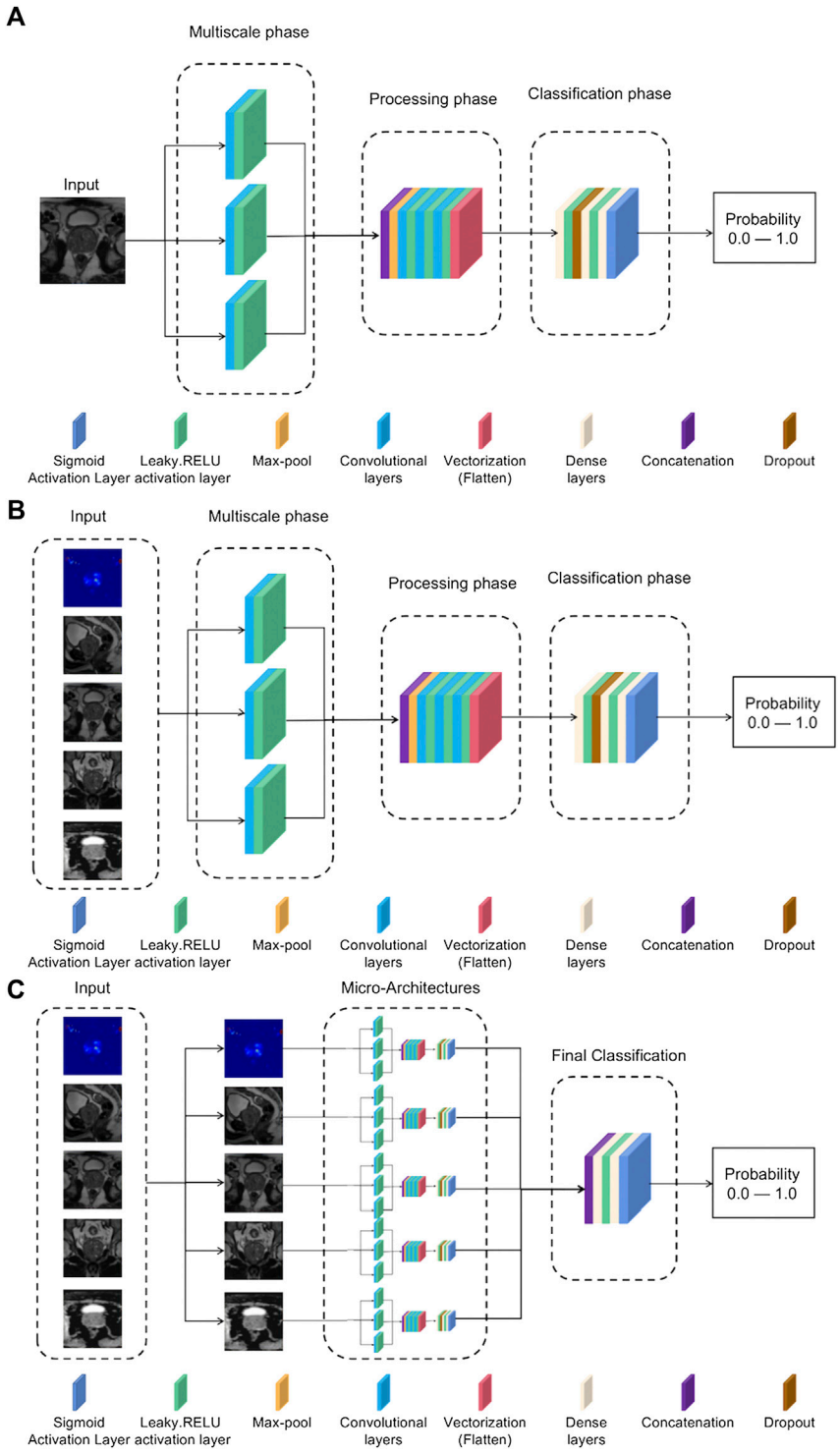


FIGURE 4
(A) Single-mode classification network structure. **(B)** Input tensor multi-modal classification network structure. **(C)** Integrated multi-modal classification network structure.

decay, LeakyReLU is still chosen as the activation function. The specific structure is shown in Figure 4A.

According to the order in the network structure, each process is introduced in turn:

1) Multiscale stage: Since tumors may be of different sizes or located in different locations, applying a series of multiscale 3D convolutions to the input data enables us to detect possible anomalies. This technique comes from the inception-v3 network, as it has been shown that multiscale filter banks can give good results on classification problems (Chollet and Ieee, 2017).
2) Processing stage: This stage starts from the max-pool layer, which allows obtaining the features of the maximum value. Convolutional filter banks are then used to obtain more complex features for further refinement of classification. This filter-based design is inspired by networks such as VGG-16 (Zhang et al., 2016). After this, a vectorized layer is used to unify all dimensions for dimensionality reduction.
3) Classification stage: The dense layer is used for classification. Experience has shown that the best training results are obtained using two dense layers with ten neurons.

2.5 Multi-modal classification network structure

Based on the single-modal classification network, we propose a multi-modal classification network structure, which tries to use multiple medical image data of patients and tries to combine the information of different attributes to achieve a better classification effect. This work proposes two different multi-modal classification network structures:

2.5.1 Input tensor multi-modal classification network structure

The goal of the multi-modal classification network structure design is to have an accurate classification effect, and it can be trained using different modes of 3D volume channels to have multiple perspectives on the diagnosis of the same lesion location. The model uses five images of the same patient as input, for which it is necessary to preselect patients with more than five images available, reducing the training set. The rest of the neural network structure is the same as the single-modal classification structure in Figure 4A, but the input consists of five images each time instead of a single image. The specific network structure is shown in Figure 4B.

2.5.2 Integrated multi-modal classification network structure

This model is the most complex in the article and is designed to use all the information previously obtained to generate a more accurate model. The network structure is based on the concept of multi-model ensemble learning (Xiao et al., 2018), which uses several lower-complexity classifiers to obtain a classifier with stronger performance. The model input uses all five types of images, but unlike the input tensor multi-modal classification structure, each type of image is now evaluated in its specific single-modal network structure, and the previously obtained weights are used to adjust the model parameters for best results. The outputs of these five sub-networks are processed in two convolutional layers, and the corresponding neurons use the leaky-RELU activation function and the Sigmoid activation function, respectively, to achieve the effect of binary classification. The specific structure is shown in Figure 4C.

2.6 Training parameters

Our model is implemented in python (version 3.8) and uses Tensorflow, Keras, OpenCV, and Cuda Toolkit as the backend DNN learning library.

We designed comparison experiments to select hyperparameters for training the classification network, including optimizers (ADAM, AdaGrad, and RMSProp), learning rate, epoch, and batch size. $1e^{-4}$ and $1e^{-5}$ learning rates were used to compare the performance of the algorithms in the comparison experiments. The model was applied to training with batches of sizes 4 and 8, while the corresponding epoch size grew from 50 to 200, increasing by 50 each time.

The parameters chosen for training the model are as follows.

1) Optimizer: The ADAM optimizer was used in this study (Zhang, 2018). The reason for choosing ADAM is that it combines the advantages of the two optimization algorithms, AdaGrad and RMSProp, and comprehensively considers the first-order moment estimation of the gradient (that is, the mean value of the gradient) and the second-order moment estimation (that is, the uncentered variance of the gradient) and calculates out the update step size. Parameter updates in ADAM are not affected by gradient scaling. Hyperparameters are well interpretable and usually require little or no tuning. At the same time, it can naturally realize the step size annealing process (automatically adjust the learning rate), which is very suitable for large-scale data and parameter scenarios such as medical image processing.
2) Batch size: Due to the small size of the database, the batch size was set to 4. This is a small-scale case and can lead to confusion in the direction of gradient descent.
3) Number of iterations: The number of iterations was set to 200, while retaining the weights of those excellent results in the validation set, thus, avoiding overfitting.
4) Learning rate: The learning rate was set to $1e^{-5}$, which is determined by the batch size.

TABLE 3 Prediction results of prostate cancer localization network.

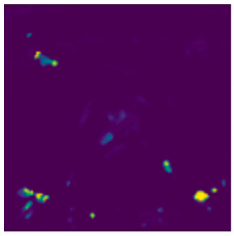
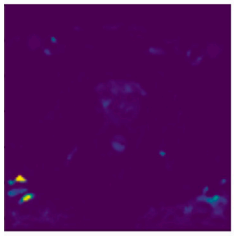
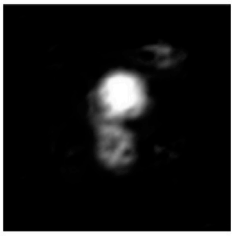
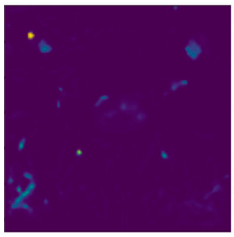
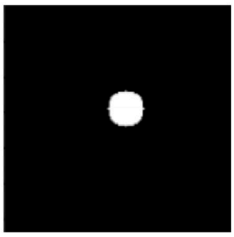
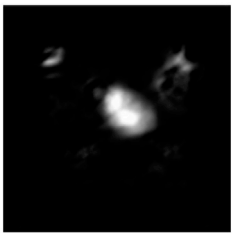
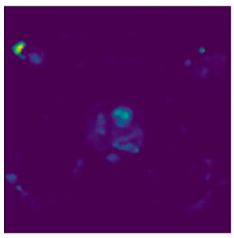
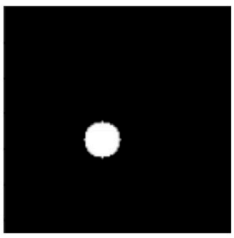
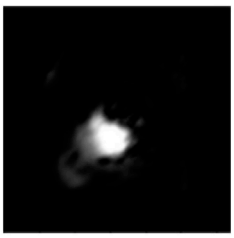
Patient ID	Two-dimensional map of interest	Labelled	Prediction	RMSE
0001				2.21
0022				1.45
0026				1.27
0057				1.61
Mean RMSE		1.64		
Color Map	-2.0 2.0			

Table 3 shows the results for four different patients in the dataset. The first column shows the patient ID., The second column shows the 2D Ktrans map, represented by a "viridis" color band for better visualization. The third and fourth columns show the two-dimensional images of the prediction results of the artificial labeling and localization network after inputting Ktrans images, which are all grayscale images, and it can be observed that the prediction results are very close to the label image. Due to the small size of the prostate, it is, on average, 40 × 30 × 20 mm. Numerically, the error between the predicted results and the labeled results for the four patients was less than 3 mm, with an average error of only 1.64 mm, and the prediction results were only about 6% error compared to the normal prostate volume. Therefore, it can be considered that the prostate cancer localization network has excellent performance and accurate prediction results, and the results can be further improved by using a larger database or better data preprocessing in the future.

5) Loss function: The loss function used in this study was focal loss, which is used to solve the problem of imbalance between positive and negative samples. The imbalance between positive and negative samples can cause the model training to fall into the local minimum of the loss function. Focal loss is used in medical image classification problems to reduce the weight of easy-to-classify samples so that the model can focus more on the hard-to-classify samples during training (Lin et al., 2020). The focal loss is calculated by multiplying the cross-entropy loss by the variable weights. Let p be the probability that the predicted sample is a positive sample ($p \in [0, 1]$) and y denote the predicted outcome ($y \in \{-1, 1\}$);

TABLE 4 Performance of prostate cancer localization network compared with previous classical segmentation methods.

Model	Sensitivity	Specificity	Jaccard index	PPV	NPV	DSC
U-Net	0.80	0.83	0.79	0.76	0.80	0.74
U-Net++	0.82	0.84	0.82	0.81	0.83	0.75
DenseNet	0.86	0.88	0.87	0.85	0.89	0.81
FCN	0.85	0.89	0.86	0.90	0.89	0.82
SegNet	0.91	0.87	0.87	0.86	0.90	0.78
Our Method	**0.92**	**0.90**	**0.89**	**0.91**	**0.93**	**0.84**

DSC, dice similarity coefficient; NPV, negative predictive value; PPV, positive predictive value. Best performance values are in bold.

then, the operation rules for the cross-entropy loss and focal loss of a sample are defined as follows:

$$CE(p, y) = \begin{cases} -\log(p), \text{if } y = 1 \\ -\log(1-p), \text{otherwise} \end{cases} \quad (1)$$

$$FL(p, y) = \begin{cases} -(1-p)^{y} \log(p), \text{if } y = 1 \\ -p^{y} \log(1-p), \text{otherwise} \end{cases} \quad (2)$$

When the sample is an easy-to-classify sample, i.e., the closer p is to 0 or 1, the smaller the calculated weight coefficient is, the smaller the proportion of the sample to the total loss, when constant; when the sample is a hard-to-classify sample, i.e., when p is close to 0.5, the larger the weight coefficient is, the larger the proportion of the sample to the total loss when *?* is constant. The focal loss used in this study makes $\gamma = 2$ to apply weights to the loss values of the hard and easy samples during the training process, making the model learning more focused on the hard-learned samples.

In the experiment, a five-fold cross-validation method was used to divide the 328 suspected lesions from the preprocessed PROSTATEx dataset into five folds according to the systematic classification to ensure that the distribution of data in each fold is consistent in terms of lesion area and benignity and malignancy, and also to avoid the problem of data leakage as much as possible. After five training sessions, the average AUC was taken as the final evaluation score.

2.7 Evaluation metrics

Network performance can be evaluated using metrics such as root mean square error (RMSE), true positive rate (TPR), true negative rate (TNR), F1-score and AUC, accuracy and confidence interval, Jaccard index, PPV, NPV, and DSC.

The RMSE is in the marked circle centroid, and the surface centroids obtained in the post-threshold prediction are calculated; RMSE is defined as follows:

$$\text{RMSE}(\mathbf{X}, h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(h(\mathbf{x}^{(i)}) - y^{(i)}\right)^2} \quad (3)$$

where RMSE (X, h) is the loss function measured in the sample set using hypothesis h, and h is the prediction function of the system, also known as the hypothesis. m is the number of instances in the dataset, $x^{(i)}$ is a vector of all eigenvalues of the ith instance in the dataset, and $y^{(i)}$ is the expected output value.

TPR, TNR, PPV, and NPV are defined as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 1 - \text{FNR} \quad (4)$$

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (7)$$

F1-score is the harmonic value of the precision and recall evaluation indicators, the best value is 1, which is defined as follows:

$$F1 - Score = 2 \cdot \frac{\text{TPR} \cdot \text{PPV}}{\text{TPR} + \text{PPV}} \quad (8)$$

AUC refers to the area under the ROC curve, which can be used to evaluate the classification quality of the classifier. The larger the value, the higher the quality of the classifier.

Accuracy is defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

The confidence in the accuracy is assessed using 95% confidence intervals, by which the range of the model's accuracy for the overall sample can be estimated, and 95% confidence intervals are defined as follows:

$$\text{P}\left(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{\text{n}}} \leq \text{M} \leq \hat{\mu} + 1.96\frac{\sigma}{\sqrt{\text{n}}}\right) \approx 0.95 \quad (10)$$

where n represents the number of selected accuracies, $\hat{\mu}$ represents the mean of all accuracies, σ represents the standard deviation of all accuracies, and M represents the desired 95% confidence interval.

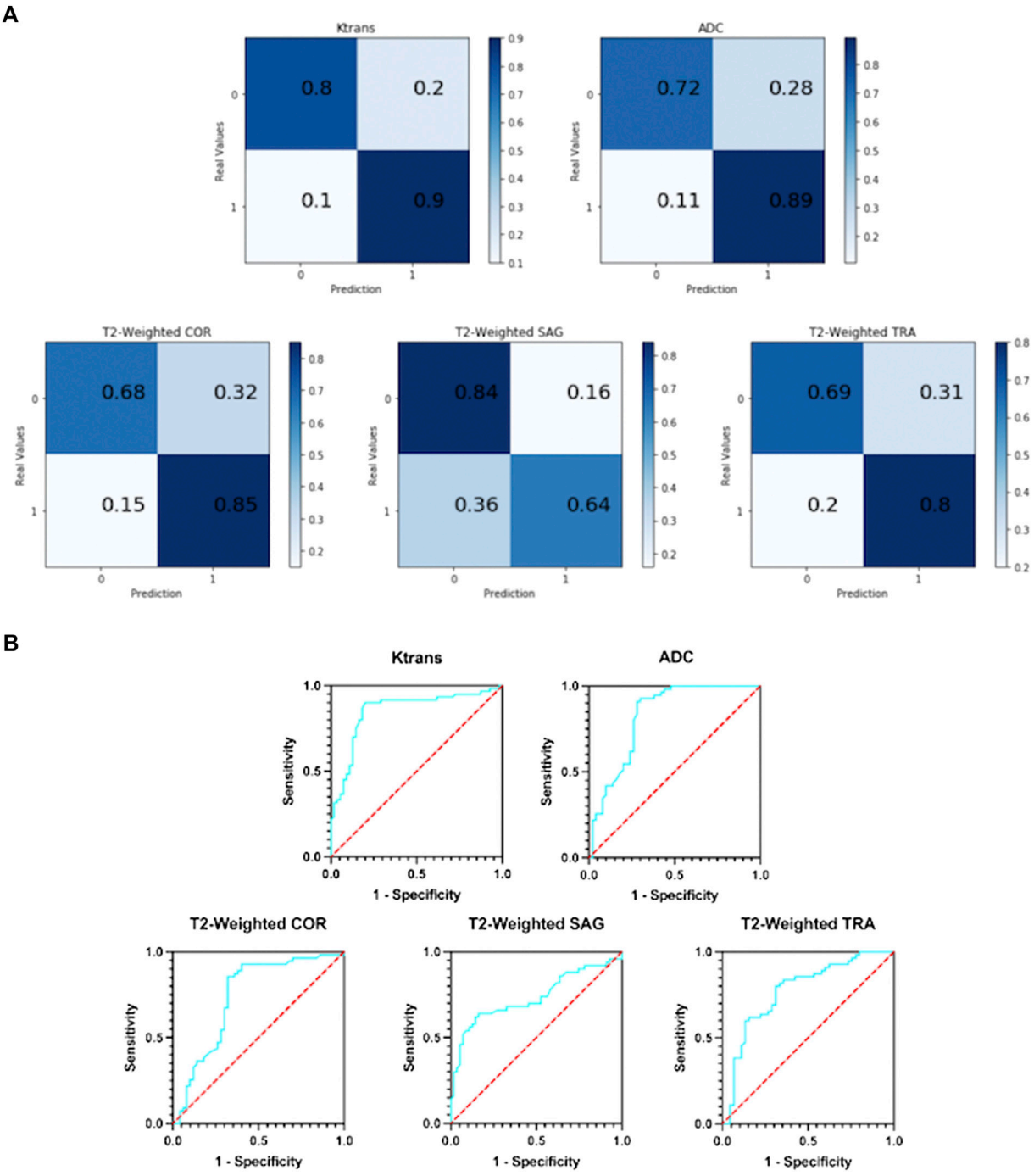


FIGURE 5
(A) Confusion matrix of five single-modal classification networks. **(B)** ROC curves of five single-mode classification networks.

Jaccard index is used to compare the similarity and difference between finite sample sets. The larger the value of the Jaccard coefficient, the higher the sample similarity. Given two sets, A and B, the Jaccard coefficient is defined as the ratio of the size of the intersection of A and B to the size of the concurrent set of A and B. It is defined as follows:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{11}$$

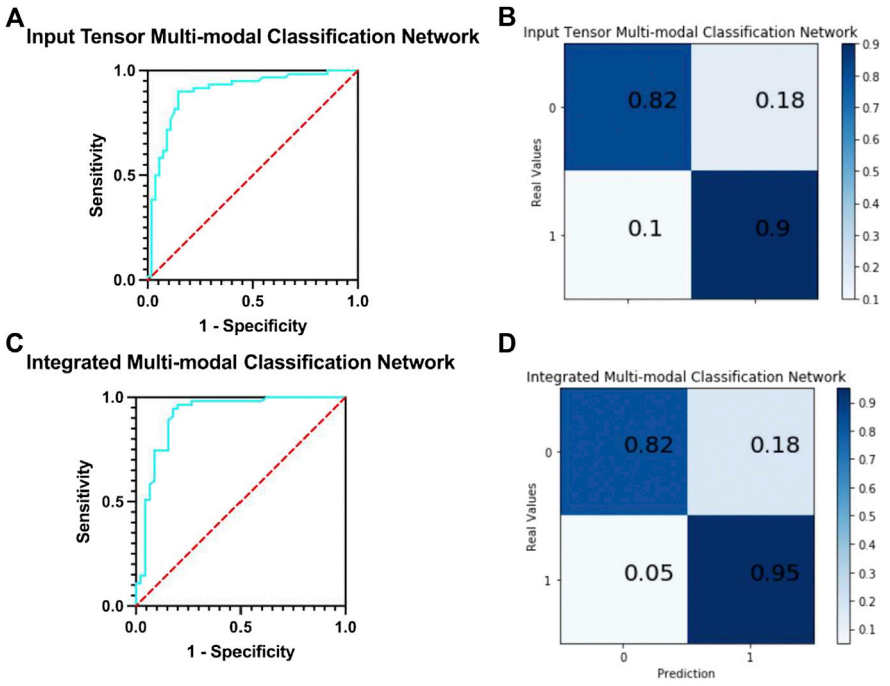


FIGURE 6
(A) ROC curve of input tensor multi-modal classification network. **(B)** Confusion matrix of input tensor multi-modal classification network. **(C)** ROC curve of integrated multi-modal classification network. **(D)** Confusion matrix of integrated multi-modal classification network.

Dice similarity coefficient (DSC) is used to measure the similarity of two sets, the value range is (0,1), and the larger the value, the more similar the two sets, commonly used in calculating the similarity of the closed region, defined as follows:

$$DSC = \frac{2TP}{FP + 2TP + FN} \tag{12}$$

3 Results

The performance of the selected different optimizers, learning rate, epoch, and batch size in the training network is shown in Supplementary Table S1. According to Supplementary Table S1, the Adam algorithm achieved the highest performance with a learning rate of $1e^{-5}$, 200 epochs, and 4 min-batches.

3.1 Prostate cancer localization network results

The results of the prostate cancer localization network are shown in Table 3. It presents experimental results from both quantitative and qualitative perspectives.

The first is the qualitative result, which visualizes the 2D portion of the 3D volume segmentation, and both the labeled images and the network predictions detect the presence of cancerous tissue. The second is the quantitative result, which expresses the root mean square error (RMSE) of each experimental image as the average RMSE in the database in millimeters while taking into account the resolution of the instrument and other issues. The use of viridis ribbons is intended to improve the readability of graphics for readers with common forms of color blindness and color vision deficiencies. Color graphics are also uniform in perception, both in regular form and when converted to black and white for printing. The performance of the prostate cancer localization network compared with previous classical segmentation methods is shown in Table 4.

As shown in the table, the prostate cancer localization network proposed in this study has improved in each index compared with previous methods, with sensitivity, specificity, Jaccard index, PPV, NPV, and DSC of 0.92, 0.90, 0.89, 0.91, 0.93, and 0.84, respectively.

Once the model is trained using multi-modal datasets, the performance of the network can be quantitatively evaluated by volumetric or regional overlapping metrics, e.g., Dice scores, as stated in the study by Yang et al. (2022). The experimental data in Table 4 fully illustrate the interpretability of the model, which

TABLE 5 Indicators of integrated multi-modal classification network, input tensor multi-modal classification network, and five single-modal classification networks.

Modality	TPR	TNR	F1-score	AUC	Accuracy
Integrated Multi-modal Classification Network	**0.95**	0.82	**0.8920**	**0.912**	**0.885**
Input Tensor Multi-modal Classification Network	0.90	0.82	0.8654	0.900	0.86
Ktrans	0.90	0.80	0.8571	0.853	0.85
ADC	0.89	0.72	0.8203	0.826	0.805
T2-Weighted COR	0.85	0.68	0.7834	0.741	0.765
T2-Weighted SAG	0.64	**0.84**	0.7636	0.735	0.74
T2-Weighted TRA	0.80	0.69	0.7583	0.775	0.745

ADC, apparent diffusion coefficient; AUC, area under curve; COR, coronal; TNR, true negative rate; TPR, true positive rate; SAG, sagittal; TRA, transverse. Best performance values are in bold.

enhances its credibility and transparency of the model, and also facilitates future improvements of the model.

3.2 Single-modal classification network results

3.2.1 Confusion matrix

Figure 5A shows the resulting confusion matrix for five different types of images as input to a single-modal classification architecture. According to Figure 5A, the Ktrans and ADC modal images perform best, and the true positive rate (TPR) and true negative rate (TNR) values are quite balanced, with an average of about 82% (85 and 80.5%, respectively).

In T2-weighted images, however, there was a large difference between the results of the three-parameter imaging. The COR image performed the best with an average of 77%, and its results were inferior to Ktrans and ADC. The mean values of SAG images and TRA images were 74 and 75%, respectively. They are not good indicators for detecting prostate cancer.

3.2.2 ROC curve

Figure 5B shows the ROC curves of the results obtained with five different types of images as input to a single-modal classification network. The ROC curve graph is a curve reflecting the relationship between sensitivity and specificity. The *X*-axis of the abscissa is 1-specificity, also known as the false positive rate (FPR), and the closer the *X*-axis is to zero, the higher the accuracy; the *Y*-axis of the ordinate is called the sensitivity, also known as the true positive rate (TPR), and the larger the *Y*-axis, the better the accuracy. So, the closer the curve is to (0,1), the better its performance.

It can be seen from Figure 5B that the performance of the Ktrans and ADC modes is relatively better. Ktrans has a smoother curve and better response, while ADC has a more abrupt response. Compared with the former two, the curves obtained by T2-weighted COR, SAG, and TRA are less suitable for the detection of prostate cancer localization.

3.2.3 Overall results

Supplementary Table S2 contains the comprehensive evaluation indicators of five single-mode classification networks, including TPR, TNR, F1-Score, AUC, and accuracy.

It can be observed that Ktrans performs the best, exceeding 0.85. ADC and COR performed slightly worse, stable at around 0.8. While SAG and TRA performed the worst, both less than 0.8. 0.8 was chosen as the threshold; based on this metric, the Ktrans modality was considered the most suitable input modality.

It can be observed that Ktrans and ADC still have good performance, reaching 85 and 83% of the area, respectively. The AUCs of the three modes of T2-weighted all fluctuate around 75%, which is not excellent.

The accuracy indicator selects 0.8 as the threshold. It was observed that among the five single-modal classification networks, Ktrans performed the best with an accuracy of 85%, followed by ADC with an accuracy of 81%.

3.3 Multi-modal classification network results

3.3.1 Input tensor multi-modal classification network results

In this network, five MRIs of the same patient with different modes are used as input. The following analyze and compare various indicators to judge whether the integration and fusion of the models can bring better performance. As shown in Figure 6A, the model obtained an AUC of 0.900, which is better than the highest value of 0.853 for the single-mode classification network. At the same time, it can be seen from the confusion matrix (Figure 6B) that the TPR and TNR values of the network are 90 and 82%, respectively, and the average value is 86%, which is slightly higher than Ktrans (average 85%), which is the best result obtained in single-mode networks and variants at present.

We compare the input tensor multi-modal classification network with five single-modal classification networks using detailed metrics in Supplementary Table S3. It can be

TABLE 6 Effect of the number of modalities on model performance.

Modality	TPR	TNR	F1-score	AUC	Accuracy
Ktrans + ADC	0.91	0.80	0.8575	0.864	0.851
Ktrans + T2-Weighted	0.89	0.81	0.8424	0.859	0.834
ADC + T2-Weighted	0.87	0.81	0.8281	0.853	0.842
Ktrans + ADC + T2-Weighted	**0.95**	**0.82**	**0.8920**	**0.912**	**0.885**

AUC, area under curve; TNR, true negative rate; TPR, true positive rate. Best performance values are in bold.

TABLE 7 Comparison between different classification networks, stratified by accuracy and 95% confidence interval.

Model	Modality	Average accuracy, 95% confidence interval
Integrated Multi-modal Classification Network	-	**0.885** [0.881, 0.889]
Input Tensor Multi-modal Classification Network	-	0.86 [0.852, 0.868]
Single-modal Classification Network	Ktrans	0.85 [0.84, 0.86]
	ADC	0.805 [0.702, 0.818]
	T2-Weighted COR	0.765 [0.75, 0.78]
	T2-Weighted SAG	0.74 [0.721, 0.759]
	T2-Weighted TRA	0.745 [0.727, 0.763]

TABLE 8 Comparison of the classification model proposed in this article with the results of previous classification models.

Model	Author	AUC
Inception V3	Quan Chen	0.83
VGG-16	Quan Chen	0.81
XmasNet	Saifeng Liu	0.84
SVM	Jarrel C.Y. Seah	0.84
3D Convolutional Neural Networks	Alireza Mehrtash	0.80
Single-modal Classification Network	-	0.853
Input Tensor Multi-modal Classification Network	-	0.900
Integrated Multi-modal Classification Network	-	0.912

observed that the input tensor multi-modal classification network outperforms the above single-modal classification network on almost all metrics, and although the lead may not be large, these small improvements play a role in the clinical detection and diagnosis of prostate cancer.

3.3.2 Integrated multi-modal classification network results

The network integrates five single-mode classification models, in each of which images of the corresponding modality of the same patient are processed. The ROC curve of the network is shown in Figure 6C. The model obtained an AUC of 0.912, which is higher than all previous models proposed in this study and has the best classification performance with a 1.2% improvement over the results of the input tensor classification network.

The confusion matrix of the integrated multi-modal classification network is shown in Figure 6D; the values of TPR and TNR are 95 and 82%, respectively, which exceed the previous best values of 90 and 82% obtained by the input tensor multi-modal classification network. This shows that the integrated multi-modal classification network, through the integration of the single-modal classification model, is not only more robust and less coupled but also optimizes the results obtained by the single-modal classification network and the input tensor multi-modal classification network to a certain extent. To test this claim, Table 5 presents the data for the remaining indicators.

Table 5 contains all the results for all the models in this study. The integrated multi-modal classification network has the optimal value for all the other indicators except TNR. For example, the prediction accuracy is improved by about 4% compared to the Ktrans single-modal classification network. Although the absolute value of the improved accuracy is not high, the higher the previous accuracy, the more significant the improvement obtained.

Meanwhile, we also conducted experiments on whether the reduction in the number of training samples would affect the classification performance of the model by setting the sample size to 50, 100, 150, and 200, respectively, and the network model was selected as the best-performing integrated multi-modal classification network in the abovementioned experiments, and the specific results are shown in Supplementary Table S4.

From Supplementary Table S4, it can be seen that the number of training samples increases from 50, 100 to 150, with the increase of training samples, the indexes have a large improvement, where the accuracy increases significantly by 20%, from 0.683 to 0.885. In the subsequent increase of training samples, from 150 to 200 to use all training samples, it can be seen that the accuracy is stable above 0.85, and the values of TPR, TNR, F1-score, and AUC indexes are stable around 0.92, 0.80, 0.86, and 0.88. It can be seen that the size of the training set has a certain influence on the performance of the classification system. With the increase of the training scale, the classification performance gradually improves, but after a certain scale, the classification performance does not change much and remains at a more stable value.

Table 6 shows the performance comparison of the integrated multi-modal classification network using some of the modalities for learning, divided into four groups for ablation experimental comparison.

Table 7 shows the comparison of the average accuracy 95% confidence intervals of the three models mentioned in the article and their different modalities. It can be observed that the average accuracy 95% confidence interval of the integrated multi-modal classification network has the least fluctuation of 0.004, which indicates that this model is more stable compared to other models.

To evaluate the classification network proposed in this article, previous networks designed using the PROSTATEx dataset were selected for comparison, as shown in Table 8.

4 Discussion

Our method successfully achieves accurate segmentation of the prostate on magnetic resonance images, and experiments with the prostate cancer localization network obtained an average root mean square error of 1.64 mm, which is approximately less than 6% error compared to the normal size of the prostate. The error of 6% is an acceptable error range, which indicates that the localization network of the prostate proposed in this study possesses a good performance. Compared with the classical medical segmentation network U-Net, the method in this study has improved by 0.12 and 0.07 in sensitivity and specificity, respectively. In terms of the Jaccard index, the performance of DenseNet, FCN, and SegNet is respectable and slightly lower than the results of this study's method by 1–3%. The prediction results of the prostate cancer localization network can be used as an evaluation index to assist radiologists in diagnosis so that doctors can locate prostate cancer more quickly and accurately.

Grand Challenges and the SPIE Medical Imaging Symposium launched an open competition in 2017 on prostate cancer prediction on magnetic resonance images to promote advances in prostate cancer detection algorithms (Litjens et al., 2014). Currently published research studies on deep learning–based prostate classification algorithms are mainly focused on PROSTATEx contestants published in *PROCEEDINGS OF SPIE*, where the use of convolutional neural networks is mostly based on the abovementioned VGG network modification. Chen et al. used a migration learning approach with Inception V3 and VGG-16, pre-trained on ImageNet, as the base network (Simonyan and Zisserman, 2014; Szegedy et al., 2016). In addition, because of the different number of positive and negative sample distributions in the cancer lesion regions, a network was trained on each region, and finally, the results of the different networks were weighted and averaged. The performance of the competition results on the PROSTATEx test set is AUC = 0.83 and AUC = 0.81, respectively (Chen et al., 2017). Liu et al. also built a new deep learning architecture, called XmasNet, based on VGG net, and obtained seven results by combining training between different sequences and calculating the weights of the seven models using a greedy algorithm, and the prediction results were taken as a weighted average, and the performance on the test set was reflected as AUC = 0.84 (Liu et al., 2017). Similar to their study, we first propose a single-modal classification network structure based on Inception-V3 and VGG-16 networks. Based on this, we further propose an input tensor multi-modal classification network structure. Combined with multi-modal ensemble learning, we propose an integrated multi-modal classification network structure. The multi-modal classification network combines the current emerging multi-modal learning and ensemble learning techniques to transfer the knowledge learned on the information-rich modality to the information-poor modality so that the learning of each modality can assist each other to achieve better classification results (Xiao et al., 2018). The integrated multi-modal classification network improved the AUC by 8.2 and 7.2% compared to Chen et al. and Liu et al.'s network, respectively, with an AUC of 0.912.

Mehrtash et al. designed a three-branch three-dimensional convolutional neural network to exploit the spatial information of the lesion and introduced regional information of the lesion location in the fully connected layer. The CNN architecture consists of three input streams: ADC map, maximum b-value from DWI, and Ktrans from DCE-MRI. Its model input is a 32*32*12 3D ROI centered on the lesion. The prediction result on the test set is AUC = 0.80 (Mehrtash et al., 2017). Unlike their design, our input stream also includes T2-weighted images, and conventional T2WI has a greater diagnostic value for prostate cancers occurring in the peripheral zone, where 70–80% of prostate cancers are clinically located (Lee et al., 2015; Israel et al., 2020). Therefore, the T2-weighted image is not only indispensable for unimodal classification networks but also has an active role in multi-modal fusion learning. Seah et al. concluded that the contrast and brightness of prostate MR images are important factors affecting the judgment of the benignity and malignancy of lesions, so they designed the auto windowing module, which can adjust the contrast and brightness of images adaptively according to the input data

and reduce the steps of image preprocessing. In addition to this, additional information such as the patient's age, the area, and the angle at which the lesion was located was used. Finally, by model integration, the network had an AUC = 0.84 on the test set (Seah et al., 2017). For the characteristics and problems of the PROSTATEx challenge dataset, we proposed image alignment, resampling, noise reduction, and normalization preprocessing methods in this study. To solve the problem of small sample data, we propose image panning, rotation, zooming, flipping, and Mixup image enhancement methods, and finally, achieve an AUC of 0.912.

As can be seen from Table 8, the AUCs of the integrated multi-modal classification network and input tensor multi-modal classification network models proposed in this study are both significantly better than the mentioned existing mainstream classification models, which are 0.912 and 0.900. Compared with the best-performing SVM and XmasNet, the integrated multi-modal classification network improves the AUC by 7.2%. Therefore, it can be proved that the proposed integrated multi-modal classification network has better classification performance.

The experimental results of our classification network show that the Ktrans modality in the single-modal classification network performs the best with an accuracy of 85%. Subsequently, by integrating and fusing different classifiers, the accuracy of the input tensor multi-modal classification model was improved to 86%. Finally, the best results are achieved in the integrated multi-modal classification model, with a small improvement of 2.5% and an accuracy of 88.5%. Therefore, we can conclude that the integration and fusion of models can lead to better performance, and the input tensor multi-modal classification network improves the performance by 1–2% compared to the single-modal classification network. On this basis, the performance of the integrated multi-modal classification network is improved by 2.5% compared to the input tensor multi-modal classification network. The successful integration of multiple models not only makes the new structure more robust and achieves the goal of low coupling but also proves that images can be combined in a decoupled manner because each single-modal classification model can be trained in a decoupled manner, and only the final network weights need to be adjusted. Huang et al. showed that the quality of the latent representation space directly determines the effectiveness of the multi-modal learning model, and the richer the variety of modalities, the more accurate the estimation of the representation space and the better the learning effect with sufficient training data (Huang et al., 2021). As can be seen from Table 6, the combination of Ktrans + ADC + T2-weighted with the highest number of modalities still achieves the best performance in all evaluation metrics, and the modal combination of Ktrans + ADC performs well in TPR, F1-Score, AUC, and accuracy, but not as well as Ktrans + T2-weighted and ADC + T2-weighted in TNR. This suggests that although T2-weighted images do not perform as well as Ktrans and ADC on single-modal classification networks, they have an active role in multi-modal fusion learning. Taking the assisted diagnosis of prostate cancer MRI as an example, multi-modal learning can aggregate information from multiple sources of data, make the representation learned by the model more complete, transfer the knowledge learned on the information-rich modality to the information-poor modality, and make the learning of each modality assist each other to achieve better classification results.

Both the prostate cancer localization network and the single-modal and multi-modal classification networks have achieved good results, but these models cannot be considered accurate enough to be used as a single diagnostic criterion. It is better suited as a support system or second opinion for radiologists, capable of detecting overlooked positive cases or speeding up the detection of possible positive cases.

Other publicly available prostate MRI datasets can be used in the future to optimize model training with the study of prostate cancer tissue contour segmentation, such as the PROMISE12 competition dataset, the main theme of which is prostate segmentation using T2WI sequences of the prostate. The data provided include 50 training samples and the corresponding prostate masks and 30 test samples. Also, in the future, when facing the multicenter prostate cancer MRI data fusion problem, it is necessary to consider the problem of certain disparity in imaging results due to scanner, parameters, and environment (Nan et al., 2022). In addition, it is possible to use a deep learning-based approach to construct scanner image invariant encoding based on the existing methods (Moyer et al., 2020). As for the interpretability of the model, in the next step, we add visual interpretation methods such as gradient interpretation method, GradCAM interpretation method, and RISE interpretation method to further solve the problem of opaque model details and achieve a "trustworthy" and "interpretable" diagnosis process.

Our study has some limitations. First, medical ethics requires that the effectiveness and safety of any new technology in the clinical application must be fully tested. Medical artificial intelligence alone has certain risks in judging diseases based on imaging data. The results of this study can only be used as a reference for radiologists' diagnoses. Second, our research is purely based on mp-MRI and does not add other types of medical indicators as parameters to the design and training of the model, such as the patient's age, weight, and PSA, to improve the generalization ability of the model. Third, in the diagnosis of prostate cancer, the DNN technology based on magnetic resonance examination is based on its database or public database and lacks external verification of a large sample size, which is also our future research direction. We look forward to developing new single-modal classification models in future work that achieve higher accuracy in the T2-weighted modality, thereby indirectly improving the performance of an integrated multi-modal classification network. Furthermore, we will cooperate with the Radiology Department of Xiangya Hospital to create our database and test our system in a real medical environment and consider inter-observer variability.

5 Conclusion

CAD of prostate cancer remains a challenging topic. In this article, we propose a localization and classification network for prostate cancer based on DNN and mp-MRI to assist radiologists in the diagnosis of such diseases. We constructed four different localization and classification networks, namely, prostate cancer localization network, single-modal classification network, input tensor multi-modal classification network, and integrated multi-modal classification network, and analyzed them in detail through experiments. The results show that the DNN-based prostate cancer localization network and integrated multi-modal classification network obtain high performance in experiments and can be used to assist radiologists in more easily localizing and classification of prostate cancer.

Data availability statement

The prostate MR images data sets generated and analyzed in this article were obtained from the PROSTATEx challenge held by Grand Challenges in conjunction with the SPIE Medical Imaging Symposium (https://PROSTATEx.grand-challenge.org/).

Ethics statement

Ethical review and approval were not required from patients due to the use of existing data from PROSTATEx Challenge.

Author contributions

ZY: project development, methodology, software, investigation, formal analysis, and writing—original draft; ZO: data curation, resources, supervision, and writing—original draft; JH: data collection, software, and validation; DQ: visualization and supervision; CQ: data curation and investigation; BO: visualization and writing—review and editing; YW: conceptualization, funding acquisition, and writing—review and editing; LW: conceptualization, resources, supervision, and administrative support.

Acknowledgments

The authors thank the PROSTATEx database for providing valuable data.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2022.918381/full#supplementary-material

References

Armato, S. G., Huisman, H., Drukker, K., Hadjiiski, L., Kirby, J. S., Petrick, N., et al. (2018). PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. *J. Med. Imaging* 5 (4), 044501. doi:10.1117/1.JMI.5.4.044501

Becker, A. S., Marcon, M., Ghafoor, S., Wurnig, M. C., Frauenfelder, T., and Boss, A. (2017). Deep learning in mammography diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. *Invest. Radiol.* 52 (7), 434–440. doi:10.1097/RLI.0000000000000358

Q. Chen, X. Xu, S. L. Hu, X. Li, Q. Zou, and Y. P. Li (Editors) (2017). "A Transfer learning approach for classification of clinical significant prostate cancers from mpMRI scans," *Conference on medical imaging - computer-aided diagnosis* (Orlando, FL. BELLINGHAM: Spie-Int Soc Optical Engineering).

F. Chollet, Ieee, Editors (2017). "Xception: Deep learning with depthwise separable convolutions," in 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, NEW YORK: IEEE).

D. Ciresan, U. Meier, and J. Schmidhuber, Ieee, Editors (2012). "Multi-column deep neural networks for image classification," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Providence, RI. NEW YORK: IEEE).

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyo, D., et al. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24 (10), 1559–1567. doi:10.1038/s41591-018-0177-5

G. E. Dahl, T. N. Sainath, and G. E. Hinton, Ieee, Editors (2013). "Improving deep neural networks for lvcsr using rectified linear units and dropout," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (CANADA, NEW YORK: IEEE Vancouver).

A. Mehrtash, A. Sedghi, M. Ghafoorian, M. Taghipour, C. M. Tempany, W. W. Wells, et al. Editors (2017). "Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks," *Conference on medical imaging - computer-aided diagnosis* (Orlando, FL. BELLINGHAM: Spie-Int Soc Optical Engineering).

J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, et al. Editors (2009). "ImageNet: A large-scale hierarchical image database," *IEEE-Computer-Society conference on computer vision and pattern recognition workshops* (Miami Beach, FL. NEW YORK: IEEE).

Huang, Y., Du, C., Xue, Z., Chen, X., Zhao, H., and Huang, L. (2021). *What makes multimodal learning better than single (provably)*. Preprints from arXiv (https://arxiv.org/pdf/2106.04538.pdf)

Israel, B., van der Leest, M., Sedelaar, M., Padhani, A. R., Zamecnik, P., and Barentsz, J. O. (2020). Multiparametric magnetic resonance imaging for the detection of clinically significant prostate cancer: What urologists need to know. Part 2: Interpretation. *Eur. Urol.* 77 (4), 469–480. doi:10.1016/j.eururo.2019.10.024

Lee, J. J., Thomas, I. C., Nolley, R., Ferrari, M., Brooks, J. D., and Leppert, J. T. (2015). Biologic differences between peripheral and transition zone prostate cancer. *Prostate* 75 (2), 183–190. doi:10.1002/pros.22903

Lin, T. Y., Goyal, P., Girshick, R., He, K. M., and Dollar, P. (2020). Focal loss for Dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (2), 318–327. doi:10.1109/TPAMI.2018.2858826

Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., and Huisman, H. (2014). Computer-aided detection of prostate cancer in MRI. *IEEE Trans. Med. Imaging* 33 (5), 1083–1092. doi:10.1109/TMI.2014.2303821

Litwin, M. S., and Tan, H. J. (2017). The diagnosis and treatment of prostate cancer: A review. *Jama* 317 (24), 2532–2542. doi:10.1001/jama.2017.7248

S. F. Liu, H. X. Zheng, Y. S. Feng, and W. Li (Editors) (2017). "Prostate cancer diagnosis using deep learning with 3D multiparametric MRI," *Conference on medical imaging - computer-aided diagnosis* (Orlando, FL. BELLINGHAM: Spie-Int Soc Optical Engineering).

F. Milletari, N. Navab, and S. A. Ahmadi, Ieee, Editors (2016). "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in 4th IEEE International Conference on 3D Vision (3DV) (Stanford, CA. NEW YORK: IEEE Stanford Univ).

Moyer, D., Ver Steeg, G., Tax, C. M., and Thompson, P. M. (2020). Scanner invariant representations for diffusion MRI harmonization. *Magn. Reson. Med.* 84 (4), 2174–2189. doi:10.1002/mrm.28243

Nan, Y., Del Ser, J., Walsh, S., Schonlieb, C., Roberts, M., Selby, I., et al. (2022). Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions. *Inf. Fusion* 82, 99–122. doi:10.1016/j.inffus.2022.01.001

Nguyen, H., Wang, Y. Z., and Wang, Z. Y. (2020). Single-shot 3D shape reconstruction using structured light and deep convolutional neural networks. *Sensors* 20 (13), E3718. doi:10.3390/s20133718

Quan, T. M., Hildebrand, D. G. C., and Jeong, W. K. (2021). FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics. *Front. Comput. Sci.* 3, 12. doi:10.3389/fcomp.2021.613981

Rueckauer, B., Lungu, I. A., Hu, Y. H., Pfeiffer, M., and Liu, S. C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. *Front. Neurosci.* 11, 682. doi:10.3389/fnins.2017.00682

J. C. Y. Seah, J. S. N. Tang, and A. Kitchen (Editors) (2017). "Detection of prostate cancer on multiparametric MRI - a ProstateX challenge runner up," *Conference on medical imaging - computer-aided diagnosis* (Orlando, FL. BELLINGHAM: Spie-Int Soc Optical Engineering).

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Preprints from arXiv (https://arxiv.org/pdf/1409.1556.pdf).

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Szegedy C., Vanhoucke V., Ioffe S., Shlens J., and Wojna Z. leee. Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA: Ieee, 2818–26.

Takeuchi, T., Hattori-Kato, M., Okuno, Y., Iwai, S., and Mikami, K. (2019). Prediction of prostate cancer by deep learning with multilayer artificial neural network. *CUAJ-Can Urol. Assoc. J.* 13 (5), E145–E150. doi:10.5489/cuaj.5526

Xiao, Y. W., Wu, J., Lin, Z. L., and Zhao, X. D. (2018). A deep learning-based multi-model ensemble method for cancer prediction. *Comput. Methods Programs Biomed.* 153, 1–9. doi:10.1016/j.cmpb.2017.09.005

Yang, G., Ye, Q. H., and Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf. Fusion* 77, 29–52. doi:10.1016/j.inffus.2021.07.016

Zhang, X. Y., Zou, J. H., He, K. M., and Sun, J. (2016). Accelerating very deep convolutional networks for classification and detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (10), 1943–1955. doi:10.1109/TPAMI.2015.2502579

Z. J. Zhang, Ieee, Editors (2018). "Improved Adam optimizer for deep neural networks," in 26th IEEE/ACM International Symposium on Quality of Service (IWQoS) (Banff, CANADA, NEW YORK: IEEE).

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 06 September 2022
DOI 10.3389/fmolb.2022.982703

Check for updates

OPEN ACCESS

EDITED BY
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

REVIEWED BY
Yujia Zhou,
Southern Medical University, China
Qian Xiaoqin,
Jiangsu University Affiliated People's Hospital, China

*CORRESPONDENCE
Jinfeng Xu,
xujinfeng@yahoo.com
Fajin Dong,
dongfajin@szhospital.com
Jun Cheng,
juncheng@szu.edu.cn

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

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 30 June 2022
ACCEPTED 26 July 2022
PUBLISHED 06 September 2022

CITATION
Zhu D, Li J, Li Y, Wu J, Zhu L, Li J, Wang Z, Xu J, Dong F and Cheng J (2022), Multimodal ultrasound fusion network for differentiating between benign and malignant solid renal tumors.
Front. Mol. Biosci. 9:982703.
doi: 10.3389/fmolb.2022.982703

COPYRIGHT
© 2022 Zhu, Li, Li, Wu, Zhu, Li, Wang, Xu, Dong and Cheng. 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.

Multimodal ultrasound fusion network for differentiating between benign and malignant solid renal tumors

Dongmei Zhu[1,2†], Junyu Li[3,4,5†], Yan Li[2], Ji Wu[6], Lin Zhu[2], Jian Li[1], Zimo Wang[1], Jinfeng Xu[1*], Fajin Dong[1*] and Jun Cheng[3,4,5*]

[1]Department of Ultrasound, The Second Clinical Medical College, Jinan University, Shenzhen, China, [2]Department of Ultrasound, The Affiliated Nanchong Central Hospital of North Sichuan Medical College, Nanchong, China, [3]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, [4]Medical Ultrasound Image Computing (MUSIC) Laboratory, Shenzhen University, Shenzhen, China, [5]Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, China, [6]Department of Urology Surgery, The Affiliated Nanchong Central Hospital of North Sichuan Medical College, Nanchong, China

Objective: We aim to establish a deep learning model called multimodal ultrasound fusion network (MUF-Net) based on gray-scale and contrast-enhanced ultrasound (CEUS) images for classifying benign and malignant solid renal tumors automatically and to compare the model's performance with the assessments by radiologists with different levels of experience.

Methods: A retrospective study included the CEUS videos of 181 patients with solid renal tumors (81 benign and 100 malignant tumors) from June 2012 to June 2021. A total of 9794 B-mode and CEUS-mode images were cropped from the CEUS videos. The MUF-Net was proposed to combine gray-scale and CEUS images to differentiate benign and malignant solid renal tumors. In this network, two independent branches were designed to extract features from each of the two modalities, and the features were fused using adaptive weights. Finally, the network output a classification score based on the fused features. The model's performance was evaluated using five-fold cross-validation and compared with the assessments of the two groups of radiologists with different levels of experience.

Results: For the discrimination between benign and malignant solid renal tumors, the junior radiologist group, senior radiologist group, and MUF-Net achieved accuracy of 70.6%, 75.7%, and 80.0%, sensitivity of 89.3%, 95.9%, and 80.4%, specificity of 58.7%, 62.9%, and 79.1%, and area under the receiver operating characteristic curve of 0.740 (95% confidence internal (CI): 0.70–0.75), 0.794 (95% CI: 0.72–0.83), and 0.877 (95% CI: 0.83–0.93), respectively.

Abbreviations: AUC: area under the receiver operating characteristic curve; CEUS: contrast-enhanced ultrasound; CT: computed tomography; ccRCC: clear cell renal cell carcinoma; chRCC: chromophobe renal cell carcinomas; MRI: magnetic resonance imaging; MUF-Net: multimodal ultrasound fusion network; NPV: negative predictive value; PPV: positive predictive value; pRCC: papillary renal cell carcinomas; ROC: receiver operating characteristic.

Conclusion: The MUF-Net model can accurately classify benign and malignant solid renal tumors and achieve better performance than senior radiologists.

Key points: The CEUS video data contain the entire tumor microcirculation perfusion characteristics. The proposed MUF-Net based on B-mode and CEUS-mode images can accurately distinguish between benign and malignant solid renal tumors with an area under the receiver operating characteristic curve of 0.877, which surpasses senior radiologists' assessments by a large margin.

KEYWORDS

renal tumor, artificial intelligence, classification, deep learning, contrast-enhanced ultrasound

Introduction

Nowadays, cancer remains a serious threat to human health worldwide. The incidence of renal cancer is increasing annually, with more than 400,000 new cases every year worldwide (Shingarev and Jaimes, 2017; Sung et al., 2021). Clear cell renal cell carcinoma (ccRCC) is the most common type of renal cell carcinoma (RCC), accounting for 80% of all RCCs. Most renal tumors do not cause obvious clinical symptoms (Ljungberg et al., 2019). About 20%~30% of patients with renal tumor resection were preoperatively misdiagnosed, resulting in unnecessary surgery with a final post-surgical diagnosis of being benign (Schachter et al., 2007). The diagnostic accuracy needs to be improved, especially for differentiating between hypoechoic benign solid tumors and malignant tumors. Noninvasive imaging modalities such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) have improved sensitivity and specificity in preoperatively differentiating among benign, malignant, and borderline tumors. Compared with these imaging modalities, contrast-enhanced ultrasound (CEUS) is more sensitive in visualizing the microcirculatory perfusion characteristics of renal tumors and thus is widely used. However, the diagnostic accuracy varies in terms of different lesion locations and radiologists.

Deep learning has shown promising results in the classification and diagnosis of renal tumors over the past few years (Oktay et al., 2018; Hussain et al., 2021; Wang et al., 2021), which does not require subjectively defined features and can capture the entirety of biological information from images compared with traditional machine learning (Sun et al., 2020; Bhandari et al., 2021; Giulietti et al., 2021; Khodabakhshi et al., 2021). The literature indicates that deep learning algorithms are better than human experts in diagnosing many kinds of diseases, such as liver, breast, lung, fundus, skin lesions (Wu et al., 2017; Lin et al., 2020; Li et al., 2021; Liu et al., 2021). These studies have shown that deep learning is stable and generalizable and can compensate for the diagnostic discrepancy among doctors with different levels of experience. To the best of our knowledge, there are no ultrasound-based radiomics studies for the differentiation between benign and malignant solid renal tumors (Esteva et al., 2017; Kokil and Sudharson, 2019; Zabihollahy et al., 2020; Hu et al., 2021; Mi et al., 2021).

In this study, we aim to establish a multimodal fusion deep neural network based on gray-scale ultrasound and CEUS images to discriminate between benign and malignant solid renal tumors. The performance of the multimodal fusion model is compared with that of the models built on single-modal data, as well as junior and senior radiologists' assessments.

Materials and methods

Patients

This retrospective study was approved by our joint institutional review boards, and anonymized data was shared through a data-sharing agreement between institutions (The Second Clinical Medical College, Jinan University, and The Affiliated Nanchong Central Hospital of North Sichuan Medical College) (No. 18PJ149 and No. 20SXQT0140). Individual consent for this retrospective analysis was waived. From June 2012 to June 2021, the information for 1547 cases was

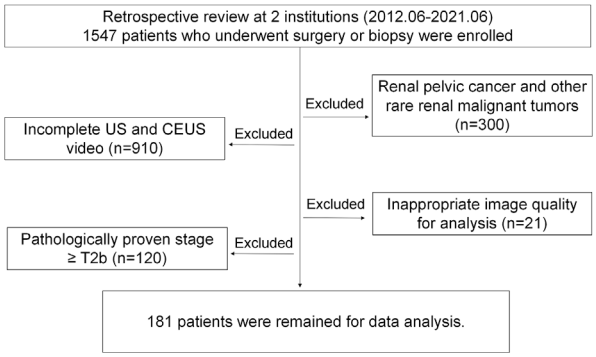


FIGURE 1
Flow diagram of patient enrollment.

TABLE 1 Patient characteristics.

	Malignant (n = 100)	Benign (n = 81)	p Value
Gender: n (%)			< 0.001*
Male	74 (74.0%)	23 (28.4%)	
Female	26 (26.0%)	58 (71.6%)	
Age: mean ± STD	58.36 ± 14.06	53.31 ± 14.00	0.017*
BMI: mean (IQR)	23.0 (22.0–25.0)	23.0 (21.0–24.0)	0.275
Tumor mean size: mean (IQR)	4.0 (3.0–6.0)	4.0 (3.0–5.0)	0.918
Clinical sign: n (%)			0.588
Waist discomfort/Fatigue	46 (46.0%)	34 (42.0%)	
No symptoms	54 (54.0%)	47 (58.0%)	
Surgery: n (%)			0.475
Partial nephrectomy	41 (41.0%)	29 (35.8%)	
Radical nephrectomy	59 (59.0%)	52 (64.2%)	

BMI, body mass index; IQR, interquartile range; STD, standard deviation.
*Statistically significant.

obtained from the surgical pathology database in the Pathology Department of The Second Clinical Medical College of Jinan University and The Affiliated Nanchong Central Hospital of North Sichuan Medical College.

The inclusion criteria were as follows: (1) preoperative CEUS examinations were performed before surgery, and (2) all cases were confirmed by pathological diagnosis after surgery. Patients were excluded based on the following criteria: (1) renal pelvis cancer and other rare types of renal malignancies; (2) patients did not receive the ultrasound and CEUS examinations; (3) poor image quality, and (4) pathologic stage ≥ T2b. Figure 1 shows the flow diagram of patient enrollment for this study. Finally, 181 patients (100 solid malignant tumors and 81 solid benign tumors) were left. Patients' demographic and clinical characteristics are shown in Table 1.

Contrast-enhanced ultrasound Imaging

CEUS examinations were performed using the following three ultrasound systems: LOGIQ E9 (GE Healthcare, Unites States), Resona7 (Mindray Ultrasound Systems, China), and IU22 (Philips Medical Systems, Netherlands), with a 1.0–5.0 MHz convex probe. CEUS was carried out by ultrasound machines with contrast-specific software and a bolus of 1.0~1.2 ml microbubble contrast agent (SonoVue; Bracco, Milan, Italy) via an antecubital vein followed by 5.0 ml normal saline with a peripheral 18~22 G needle. The CEUS digital video was at least 3~5 min long each time. During contrast-enhanced imaging, low-acoustic power modes were used with a mechanical index of 0.05~0.11. All the CEUS examination videos were retrospectively analyzed by two groups of radiologists with different levels of experience (three junior radiologists with more than 5~6 years of experience in CEUS imaging and three senior radiologists with more than 10~15 years of experience in CEUS imaging).

In this study, we used the following phase terms: (1) cortical phase began 10~15 s after injection, and (2) medullary phase approximately began 30~45 s after injection until the microbubble echoes disappeared. The entire course of CEUS was saved as Digital Imaging and Communication in Medicine format.

Data annotation and preprocessing

CEUS videos were annotated using the Pair annotation software package (Liang et al., 2022; Qian et al., 2022). In each CEUS video, about 50~60 images were selected from the cortical and medullary phases. A senior radiologist classified the tumor as either benign or malignant and annotated its location in each selected image by a bounding box. Then, according to the bounding boxes, these images were cropped into smaller images as region of interest to exclude the non-tumor regions (Figure 2). This resulted in a total of 9794 images, of which 3659 images were benign (including 1531 from 36 atypical benign cases and 2128 images from 45 typical benign cases and 6135 images were malignant (including 2964 images from 62 ccRCC cases; 2114 images from 25 pRCC cases; 1057 images from 13 chRCC cases) (Table 2).

Multimodal ultrasound fusion network

The dataset used in this study contained B-mode and CEUS-mode images, and they were in one-to-one correspondence. Therefore, we proposed the MUF-Net to take full advantage of the multimodal features, which can independently extract features from each of the two modalities and learn adaptive weights to fuse features for each sample.

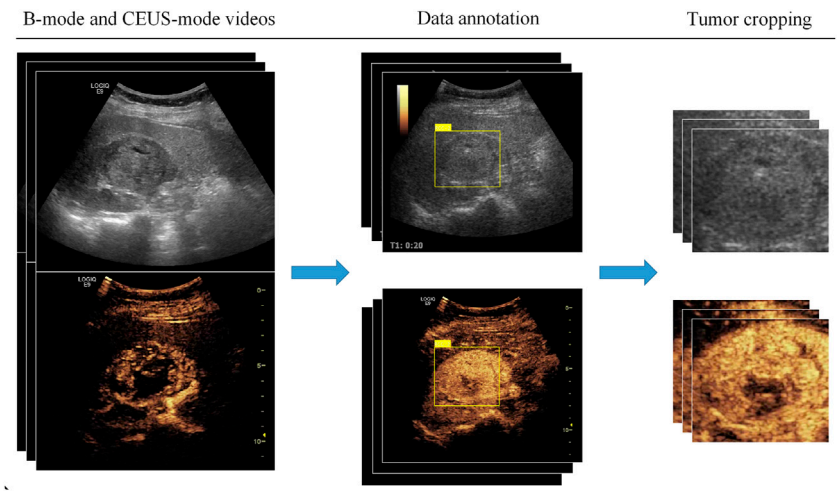


FIGURE 2
Data annotation and preprocessing.

TABLE 2 Number distribution of patients and images among histologic types.

	Benign			Malignant			
	Total	Atypical	Typical	Total	ccRCC	pRCC	chRCC
Patients	81	36	45	100	62	25	13
Images	3659	1531	2128	6135	2964	2114	1057

ccRCC, clear cell Renal cell carcinoma; chRCC, chromophobe renal cell carcinomas; pRCC, papillary renal cell carcinomas.

The overall architecture of MUF-Net is shown in Figure 3, I_{B-mode} and $I_{CEUS-mode}$ represented the B-mode and CEUS-mode images, respectively. We used two independent EffecientNet-b3 as the backbone to extract features from B-mode and CEUS-mode images. The backbone had an input size of 300 × 300 × 3 and an output size of 10 × 10 × 1536 after five down sampling blocks. To reduce the parameters of the network and prevent overfitting, we used a global average pooling layer to downsample the output feature maps of each backbone from 10 × 10 × 1536 to 1 × 1 × 1536. Subsequently, we fused the features of the two modalities. Considering that the features of the two modalities in each sample may contribute differently to the final prediction, we designed two attention blocks sharing weights to produce adaptive weights α and β for modality fusion. The feature maps of the two modalities were subsequently weighted and summed based on the adaptive weights, yielding a fused feature map of 1 × 1 × 1536. Finally, through a fully connected layer and a softmax layer, the classification result was given.

Notably, to improve the feature learning ability of each backbone, we added two classifiers C_{B-mode} and $C_{CEUS-mode}$ for each single modality, as shown in Figure 3, which independently took B-mode features and CEUS-mode features as input and calculated the loss of each mode (L_{B-mode} and $L_{CEUS-mode}$), respectively. The total loss was defined by Eq. 1. The two losses, L_{B-mode} and $L_{CEUS-mode}$, were only used during training, and the final classification result was given by the multimodal classifier, C_{B+CEUS}. Due to this reason, the multimodal loss, L_{B+CEUS}, had a higher weight than the other two losses. For the calculation of the three losses in Eq. 1, we employed the class-balanced focal loss (Zhang et al., 2021).

$$L_{total} = 3L_{B+CEUS} + L_{B-mode} + L_{CEUS-mode} \quad (1)$$

Implementation details

All experiments were conducted using five-fold cross-validation. For data splitting, we ensured that the images from the same patient went into either the training set or the test set to avoid the data leakage problem. To avoid model overfitting, data

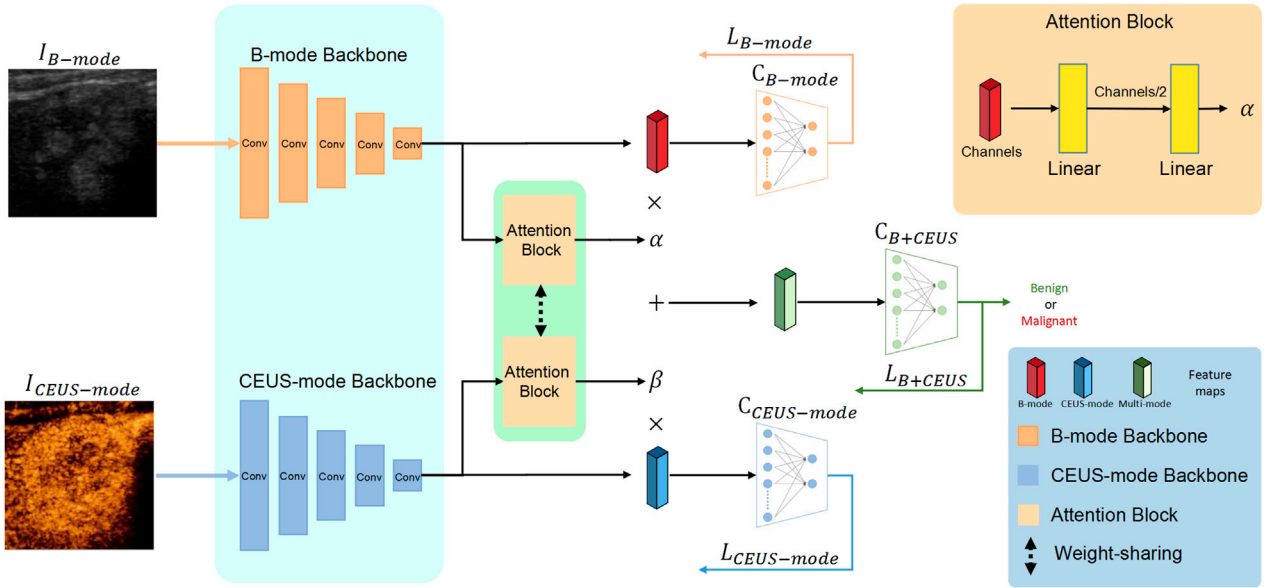


FIGURE 3
Overall architecture of the proposed MUF-Net framework.

augmentation techniques were applied to the training set, which included random spatial transformations, random non-rigid body transformations, and random noise.

The dataset used in this study had a weak class imbalance problem. The ratio of benign images to malignant images was 3:5. Re-sampling techniques (Zhang et al., 2021) were popularly used for dealing with long-tailed problems. We used class-balanced sampling to alleviate class imbalance by first sampling a class and then selecting an instance from the chosen class (Kang et al., 2019).

All backbones used in these experiments were pretrained on ImageNet. All models used in the experiments were implemented by PyTorch on a NVIDIA 3090 GPU. The stochastic gradient descent optimizer was used with a learning rate of 0.05 which was halved every 10 epochs. In each round of five-fold cross-validation, models based on B-mode, CEUS-mode, and B + CEUS mode were trained for 100 epochs, respectively, and the models with the highest accuracy on the validation set were saved.

Radiologists' assessments

Original uncropped CEUS videos and images were evaluated by three junior and three senior radiologists and manually classified as benign or malignant. The radiologists were blinded to any clinical information of the patients. Intraclass correlation coefficients (ICCs) were used to evaluate the inter-rater agreement within each radiologist group, with an ICC greater than 0.75 indicating good agreement.

Statistical analysis

All statistical analyses were performed using the SciPy package in Python (version 3.8). Depending on whether data conformed to a normal distribution, continuous variables were compared using the Student's t-test or the Mann-Whitney U test. The non-ordered categorical variables were compared by the chi-square test. Receiver operating characteristic (ROC) curve analysis was used to evaluate the performance of junior radiologists, senior radiologists, individual modality-based networks, and MUF-Net. In addition, we also used other metrics to evaluate model performance from various aspects, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Comparison of the difference between areas under the ROC curve (AUCs) was performed using the Delong test. A two-sided p value < 0.05 was considered statistically significant.

Results

Patient characteristics

The age of the patients in the benign tumor group was less than that of the patients in the malignant tumor group (58.36 ±

TABLE 3 Classification performance of deep learning models and radiologists.

	AUC (95% CI)	Accuracy (%)	Sensitivity (%	Specificity (%)	PPV (%)	NPV (%)
Junior radiologists	0.740 (0.70–0.75)	70.6	89.3	58.7	58.0	89.5
Senior radiologists	0.794 (0.72–0.83)	75.7	95.9	62.9	62.3	95.9
B-mode-Net	0.820 (0.70–0.83)	74.5	75.0	77.0	73.4	62.3
CEUS-mode-Net	0.815 (0.75–0.89)	73.9	73.8	73.2	72.5	62.2
MUF-Net	0.877 (0.83–0.93)	80.0	80.4	79.1	86.9	70.0

CI, confidence interval; CEUS-mode, contrast-enhanced ultrasound mode; MUF-Net, multimodal ultrasound fusion network; AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.

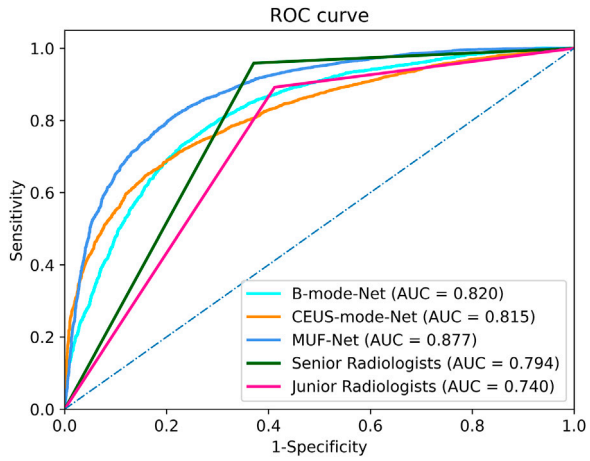


FIGURE 4
The receiver operating characteristic curves of the MUF-Net, single-mode models, and radiologists' assessments in the test cohort.

14.06 vs. 53.31 ± 14.00 years) (p = 0.017). Regarding gender distribution, there was a significant difference between these two groups (p < 0.001), with male patients being more frequent in the malignant group than in the benign group. Patient characteristics are shown in Table 1.

Performance of radiologists' assessments

The ICCs in the junior and senior radiologist groups were 0.81 and 0.83, respectively, indicating good inter-rater agreement. Each radiologist classified a tumor as benign or malignant, and the radiologists' assessments in each group were merged by majority voting. The performance of radiologists' assessments is shown in Table 3. The AUC, accuracy, sensitivity, specificity, PPV, and NPV of junior radiologists were 0.740 (95% confidence interval (CI): 0.70–0.75), 70.6%, 89.3%, 58.7%, 58.0%, and 89.5%, respectively. The AUC, accuracy, sensitivity, specificity, PPV, and NPV of senior radiologists were 0.794 (95% CI: 0.72–0.83), 75.7%, 95.9%, 62.9%, 62.3%, and 95.9%, respectively. The ROC curves for the test set were shown in Figure 4.

Performance of deep learning models

The AUC, accuracy, sensitivity, specificity, PPV, and NPV of EffecientNet-b3 network trained on B-mode images were 0.820 (95% CI: 0.72–0.83), 74.5%, 75.0%, 77.0%, 73.4%, and 62.3%, respectively (Table 3). The AUC, accuracy, sensitivity, specificity, PPV, and NPV of EffecientNet-b3 network trained on CEUS-mode images were 0.815 (95% CI: 0.75–0.89), 73.9%, 73.8%, 73.2%, 72.5%, and 62.2%, respectively. The AUC, accuracy, sensitivity, specificity, PPV, and NPV of MUF-Net trained on B-mode and CEUS-mode images were 0.877 (95% CI: 0.83–0.93), 80.0%, 80.4%, 79.1%, 86.9%, and 70.0%, respectively. The proposed MUF-Net significantly outperformed junior and senior radiologists (p < 0.001).

Discussion

This study explored the performance of deep learning based on ultrasound images for benign/malignant classification of solid renal tumors. We proposed the MUF-Net for fusing complementary features of two modalities, which used two independent EffecientNet-b3 as backbones to extract features from B-mode and CEUS-mode ultrasound images. Our method reached expert-level diagnostic performance and had a higher diagnostic PPV compared with radiologists.

CEUS is an important supplement to conventional ultrasound, CT, and MRI in diagnosing solid renal tumors. Compared with conventional ultrasound, CEUS can display perfusion characteristics of renal tumors in cortical and medullary phases in real-time, which is an important tool for improving differential diagnosis of benign and malignant renal tumors. In this study, we observed that the deep learning models built on either B-mode or CEUS-mode images achieved better performance than junior or senior radiologists. Moreover, the

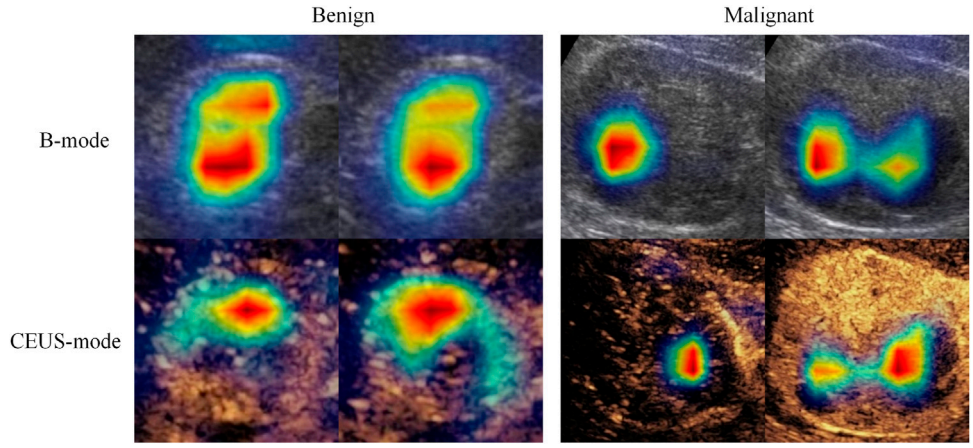


FIGURE 5
Feature heatmaps of a benign tumor and a malignant tumor to show B-mode and CEUS-mode images contain complementary information for diagnosis. The red color represents higher weights (i.e., the network pays more attention to this region).

deep learning model combing B-mode and CEUS-mode images further improved the classification performance. This indicates that B-mode and CEUS-mode images have complementary information for diagnosing solid renal tumors. To verify this point, we used class activation maps (Xu et al., 2022) to visualize the important regions that the model paid attention to in B-mode and CEUS-mode images. As shown in Figure 5, the important regions contributing to the final prediction were different between B-mode and CEUS-mode images. In other words, the MUF-Net can automatically extract complementary features from different modalities to improve the classification performance.

According to the results from the comparative experiments, we found that the performance was similar between the two models built on B-mode or CEUS-mode images. The B-mode images-based model had slightly better performance, which might be due to the different microcirculatory perfusion characteristics of solid renal tumors. Solid renal tumors of different histopathological types have different vascular density, fat content, blood flow velocity, and the severity of arteriovenous fistulas.

Lin et al. reported an AUC of 0.846 for the classification of benign and malignant renal tumors on enhanced CT images using inception-v3 (Lin et al., 2020). Xu et al. used ResNet-18 to classify multimodal MRI images of renal tumors, with AUCs of 0.906 and 0.846 on T2WI and DWI, respectively (Xu et al., 2022). The AUC was improved to 0.925 by fusing the two modalities, exceeding the diagnostic performance of highly qualified radiologists. The results of this study were similar. The MUF-Net based on multimodal data surpassed the models based on individual modalities by a large margin. Therefore, we inferred that the adaptive weights learned by MUF-Net could help the network acquire the complementary information from the two modalities to improve the classification performance.

This study had several limitations. First, the classical and well-established convolutional neural network, EffecientNet-b3, was selected as the backbone based on previous experiments, which may not be optimal. The characteristics of multimodal ultrasound imaging data need to be analyzed in-depth, and other deep neural networks will be attempted in the future to see if better performance can be achieved. Second, only images of tumor regions were cropped and used in data analyses, and regions of tumor periphery might be able to provide more information to improve model performance, which requires further experimental analyses. Third, our model was implemented using the ultrasound images collected from two hospitals only. A larger dataset acquired from more hospitals with different types or models of ultrasound equipment may have the potential to further improve the performance and generalization ability of our model.

Conclusion

In this study, the proposed MUF-Net is able to classify benign and malignant solid renal tumors accurately, by extracting complementary features from B-mode and CEUS-mode images, which outperforms senior radiologists by a large margin.

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.

Ethics statement

The studies involving human participants were reviewed and approved by the studies were reviewed and approved by The Second Clinical Medical College, Jinan University, and The Affiliated Nanchong Central Hospital of North Sichuan Medical College. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

DZ and JC contributed to conception and design of the study. DZ, YL, JW, LZ, JiL, and ZW collected the data. DZ, JuL, and JC analyzed and interpreted the data. DZ and JL wrote the first draft of the manuscript. JX, FD, and JC supervised the project and revised the manuscript. All authors contributed to the manuscript and approved the submitted version.

Funding

This project was supported by the Scientific Research Program of the Sichuan Health Commission and the Commission of Science and Technology of Nanchong (No. 21PJ195 and No. 20SXQT0140), National Natural Science Foundation of China (No. 61901275), and Shenzhen University Startup Fund (No. 2019131).

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

Bhandari, A., Ibrahim, M., Sharma, C., Liong, R., Gustafson, S., and Prior, M. (2021). CT-Based radiomics for differentiating renal tumours: A systematic review. *Abdom. Radiol.* 46 (5), 2052–2063. doi:10.1007/s00261-020-02832-9

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542 (7639), 115–118. doi:10.1038/nature21056

Giulietti, M., Cecati, M., Sabanovic, B., Scirè, A., Cimadamore, A., Santoni, M., et al. (2021). The role of artificial intelligence in the diagnosis and prognosis of renal cell tumors. *Diagn. (Basel, Switz.* 11 (2), 206. doi:10.3390/diagnostics11020206

Hu, H.-T., Wang, W., Chen, L.-D., Ruan, S.-M., Chen, S.-L., Li, X., et al. (2021). Artificial intelligence assists identifying malignant versus benign liver lesions using contrast-enhanced ultrasound. *J. Gastroenterol. Hepatol.* 36 (10), 2875–2883. doi:10.1111/jgh.15522

Hussain, M. A., Hamarneh, G., and Garbi, R. (2021). Learnable image histograms-based deep radiomics for renal cell carcinoma grading and staging. *Comput. Med. Imaging Graph.* 90, 101924. doi:10.1016/j.compmedimag.2021.101924

Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., et al. (2019). *Decoupling representation and classifier for long-tailed recognition*. Available at: http:/arXiv.org/abs/1910.09217. doi:10.48550/arXiv.1910.09217

Khodabakhshi, Z., Amini, M., Mostafaei, S., Haddadi Avval, A., Nazari, M., Oveisi, M., et al. (2021). Overall survival prediction in renal cell carcinoma patients using computed tomography radiomic and clinical information. *J. Digit. Imaging* 34 (5), 1086–1098. doi:10.1007/s10278-021-00500-y

Li, W., Lv, X.-Z., Zheng, X., Ruan, S.-M., Hu, H.-T., Chen, L.-D., et al. (2021). Machine learning-based ultrasomics improves the diagnostic performance in differentiating focal nodular hyperplasia and atypical hepatocellular carcinoma. *Front. Oncol.* 11, 544979. doi:10.3389/fonc.2021.544979

Liang, J., Yang, X., Huang, Y., Li, H., He, S., Hu, X., et al. (2022). Sketch guided and progressive growing GAN for realistic and editable ultrasound image synthesis. *Med. Image Anal.* 79, 102461. doi:10.1016/j.media.2022.102461

Lin, Y., Feng, Z., and Jiang, G. (2020). Gray-scale ultrasound-based radiomics in distinguishing hepatocellular carcinoma from intrahepatic mass-forming holangiocarcinoma. *Chin. J. Med. Imaging* 28 (4), 269–272. doi:10.3969/j.issn.1005-5185.2020.04.007

Liu, Y., Zhou, Y., Xu, J., Luo, H., Zhu, Y., Zeng, X., et al. (2021). Ultrasound molecular imaging-guided tumor gene therapy through dual-targeted cationic microbubbles. *Biomater. Sci.* 9 (7), 2454–2466. doi:10.1039/d0bm01857k

Ljungberg, B., Albiges, L., Abu-Ghanem, Y., Bensalah, K., Dabestani, S., Fernández-Pello, S., et al. (2019). European association of urology guidelines on renal cell carcinoma: The 2019 update. *Eur. Urol.* 75 (5), 799–810. doi:10.1016/j.eururo.2019.02.011

Mi, S., Bao, Q., Wei, Z., Xu, F., and Yang, W. (2021). "*MBFF-Net: Multi-Branch feature fusion network for carotid plaque segmentation in ultrasound*," in International conference on medical image computing and computer-assisted intervention. (Strasbourg, France. Springer, Cham), 313–322.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). *Attention u-net: Learning where to look for the pancreas*. Available at: http:/arXiv.org/abs/1804.03999.doi:10.48550/arXiv.1804.03999

Qian, J., Li, R., Yang, X., Huang, Y., Luo, M., Lin, Z., et al. (2022). Hasa: Hybrid architecture search with aggregation strategy for echinococcosis classification and ovary segmentation in ultrasound images. *Expert Syst. Appl.* 202, 117242. doi:10.1016/j.eswa.2022.117242

Schachter, L. R., Cookson, M. S., Chang, S. S., Smith, J. A., Dietrich, M. S., Jayaram, G., et al. (2007). Second prize: Frequency of benign renal cortical tumors and histologic subtypes based on size in a contemporary series: What to tell our patients. *J. Endourol.* 21 (8), 819–823. doi:10.1089/end.2006.9937

Shingarev, R., and Jaimes, E. A. (2017). Renal cell carcinoma: New insights and challenges for a clinician scientist. *Am. J. Physiol. Ren. Physiol.* 313 (2), F145–F154. doi:10.1152/ajprenal.00480.2016

Kokil, P., and Sudharson, S. (2019). Automatic detection of renal abnormalities by off-the-shelf CNN features. IETE J. Edu. 60 (1), 14–23. doi:10.1080/09747338.2019.1613936

Sun, X.-Y., Feng, Q.-X., Xu, X., Zhang, J., Zhu, F.-P., Yang, Y.-H., et al. (2020). Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: Comparison with expert-level radiologists. *AJR. Am. J. Roentgenol.* 214 (1), W44–W54. doi:10.2214/AJR.19.21617

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., and Li, J. (2021). "*TransBTS: Multimodal brain tumor segmentation using transformer*," in International conference on medical image computing and computer-assisted intervention. (Strasbourg, France. Springer, Cham), 109–119.

Wu, J., You, L., Lan, L., Lee, H. J., Chaudhry, S. T., Li, R., et al. (2017). Semiconducting polymer nanoparticles for centimeters-deep photoacoustic imaging in the Second near-infrared window. *Adv. Mat.* 29, 1703403. doi:10.1002/adma.201703403

Xu, Q., Zhu, Q., Liu, H., Chang, L., Duan, S., Dou, W., et al. (2022). Differentiating benign from malignant renal tumors using T2- and diffusion-weighted images: A comparison of deep learning and radiomics models versus assessment from radiologists. *J. Magn. Reson. Imaging.* 55 (4), 1251–1259. doi:10.1002/jmri.27900

Zabihollahy, F., Schieda, N., Krishna, S., and Ukwatta, E. (2020). Automated classification of solid renal masses on contrast-enhanced computed tomography images using convolutional neural network with decision fusion. *Eur. Radiol.* 30 (9), 5183–5190. doi:10.1007/s00330-020-06787-9

Zhang, Y., Wei, X.-S., Zhou, B., and Wu, J. (2021). Bag of tricks for long-tailed visual recognition with deep convolutional neural networks. *Proc. AAAI Conf. Artif. Intell.* 35 (4), 3447–3455.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 September 2022
DOI 10.3389/fgene.2022.951077

Check for updates

OPEN ACCESS

EDITED BY
William C. Cho,
QEH, Hong Kong SAR, China

REVIEWED BY
Feng Ren,
Xinxiang Medical University, China
Ze-Kun Liu,
Fourth Military Medical University, China
Haijiao Yan,
Department of Oncology, First Affiliated Hospital of Soochow University, China

*CORRESPONDENCE
Yanyan Liu,
liuyanyan725@163.com
Yufeng Gao,
aygyf@ahmu.edu.cn
Honghai Xu,
xuhonghaibl@163.com

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

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 16 June 2022
ACCEPTED 23 August 2022
PUBLISHED 09 September 2022

CITATION
Qin J, Sun W, Zhang H, Wu Z, Shen J, Wang W, Wei Y, Liu Y, Gao Y and Xu H (2022), Prognostic value of LECT2 and relevance to immune infiltration in hepatocellular carcinoma.
Front. Genet. 13:951077.
doi: 10.3389/fgene.2022.951077

COPYRIGHT
© 2022 Qin, Sun, Zhang, Wu, Shen, Wang, Wei, Liu, Gao 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.

Prognostic value of LECT2 and relevance to immune infiltration in hepatocellular carcinoma

Jiangfeng Qin[1,2†], Weijie Sun[2†], Hui Zhang[1†], Zihao Wu[3], Jiapei Shen[2], Wenhai Wang[1], Yuanyuan Wei[4,5], Yanyan Liu[5*], Yufeng Gao[2,5*] and Honghai Xu[1,5*]

[1]Department of Pathology, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [2]Department of Infectious Diseases, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [3]Department of Pathology, the Fourth Affiliated Hospital of Anhui Medical University, Hefei, China, [4]Department of Hospital Infection Prevention and Control, the First Affiliated Hospital of Anhui Medical University, Hefei, China, [5]Anhui Center for Surveillance of Bacterial Resistance, Hefei, China

Background: Previous studies have shown that Leukocyte cell-derived chemotaxin2 (LECT2) is associated with the development of HCC. However, there are still no studies with a comprehensive analysis of the role of LECT2 in hepatocellular carcinoma (HCC).

Methods: TCGA data sets were used to analyze the expression of LECT2 in HCC. In addition, the prognostic value of LECT2 in HCC was also investigated. DriverDBv3 was used to analyze the Mutation, CNV, and methylation profiles of LECT2. And, validated by immunohistochemistry in 72 HCC samples. The prognostic value of LECT2 and the correlation with clinicopathological features were analyzed. The GO/KEGG enrichment analysis of LECT2 co-expression and gene set enrichment analysis (GSEA) was performed using the R software package. The PPI interaction network was constructed by Search Tool for the Retrieval of Interacting Genes (STRING) database. Immune infiltration was estimated by the XCELL, TIMER, QUANTISEQ, MCPCOUNTER, EPIC, CIBERSORT abs and CIBERSORT algorithms, and Spearman was used to analyzing their correlation with LECT2. Moreover, we analyzed the correlation of LECT2 expression with immune checkpoint molecules and HLA genes. Finally, we analyzed the IC50 values of six chemotherapeutic drugs by the pRRophetic package.

Results: Reduced LECT2 expression levels found in HCC patients. Moreover, decreased levels of LECT2 were associated with poor overall survival, disease-free survival, disease-specific survival, and progression-free survival. Besides, methylation was significantly associated with LECT2 expression. The functional enrichment analysis revealed that LECT2 may affect HCC progression through various pathways such as JAK/STAT signaling pathway, cell cycle, and pathways in cancer. Additionally, the results showed that LECT2 expression was negatively correlated with immune infiltration of B cells, Neutrophil, Monocyte, Cancer-associated fibroblast, and Myeloid dendritic cell, and positively correlated with T cell $CD8^+$ naive, Endothelial cell, and Hematopoietic stem cell. LECT2 expression was negatively correlated with multiple immune checkpoint molecules and HLA genes. Chemosensitivity analysis showed that chemosensitivity was lower in the LECT2 high expression group. We

validated the prognostic value of LECT2 and analysis of clinicopathological features showed a lower TNM stage in the group with high expression of LECT2.

Conclusion: Low expression of LECT2 in HCC is closely associated with poor prognosis, LECT2 may have potential clinical applications due to its unique immunological effects.

KEYWORDS

leukocyte cell-derived chemotaxin2, HCC, immune, bioinformatics analysis, prognosis

Introduction

Cancer is the leading cause of death in most countries of the world in the 21st century, and the increase in cancer incidence and mortality has caused widespread concern worldwide (Sung et al., 2021). Immunotherapy is one of the breakthroughs in cancer treatment and is becoming increasingly popular (Zhang and Zhang, 2020). However, a large number of cancer patients still have a poor prognosis due to distant metastases and cancer recurrence, with 5-years survival rates below 20% for many cancer subtypes (Bengtsson et al., 2020). Therefore, it is crucial to discover meaningful biomarkers to assess the prognosis of cancer.

Leukocyte cell-derived chemotaxin2 (LECT2) is a 16-kDa secreted protein (Ito et al., 2003), LECT2 was first reported as a chemotactic factor to promote the migration of neutrophils (Yamagoe et al., 1996). Recent evidence suggests that LECT2 is strongly associated with multiple disease progression, including renal amyloidosis (Comenzo, 2014), diabetes (Lan et al., 2014), and sepsis (Ando et al., 2012). In addition, identified as a hormone-like hepatokine, LECT2 is highly expressed in the liver (Zhu et al., 2022). Therefore, LECT2 is also closely associated with a variety of liver diseases, such as non-alcoholic fatty liver disease (NAFLD) (Yoo et al., 2017), and liver fibrogenesis (Xu et al., 2019), and hepatocellular carcinoma (Chen et al., 2016). We recently reported that LECT2 can suppress the migration and tube formations of endothelial cells via binding to Tie1 (Xu et al., 2019). Loss of LECT2 results in an increase of $CD4^+$ T cells in the spleen (Greenow et al., 2018). By activating LPS signaling in macrophages, LECT2 links obesity to hepatic inflammation (Takata et al., 2021). All of this evidence suggests that LECT2 is closely associated with immune cell infiltration and may serve as a promising target for cancer immunotherapy.

Our study found that the expression level of LECT2 in HCC correlated with prognosis. According to the report, HCC progression is inhibited by LECT2 by controlling inflammatory monocytes (L'Hermitte et al., 2019). However, there is a lack of comprehensive studies on the prognostic value and the role of LECT2 in HCC in terms of immunotherapy. In this study, we investigated the expression and prognostic value of LECT2 in HCC. We also searched for possible signaling pathways by which LECT2 affects HCC and focused on exploring the correlation between LECT2 and immune infiltration. In addition to this, we performed a comprehensive analysis of the clinicopathological information of LECT2 in HCC patients. To our knowledge, this is the first study to analyze in detail the role of LECT2 in HCC, providing a reference for the use of LECT2 in HCC patients.

Materials and methods

Data download

The mRNA expression profile data from 33 different cancer patients from The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga/), and missing data were removed information.

Expression and survival analysis of LECT2

Using the limma package, we analyzed differential gene expression in 33 cancers and finally identified LECT2 with high expression levels in HCC and CHOL. Survival curves of high and low LECT2 expression level groups in HCC and CHOL were plotted using the Kaplan-Meier method to determine their prognostic value, with prognostic endpoints including overall survival (OS), disease-free survival (DFS), disease-specific survival (DSS) and progression-free survival (PFS).

Exploring gene mutation, CNV (copy number variations), and methylation spectrum of LECT2

We explored the LECT2 gene mutation, CNV, and methylation spectrum using DriverDBv3 (http://driverdb.tms.cmu.edu.tw/). These data are developed and obtained for free.

Enrichment analysis of LECT2 and PPI analysis

LinkedOmics database (http://www.linkedomics.org/login.php) is a fully functional multi-omics database that can be

used for association analysis between genes (Vasaikar et al., 2018). LECT2 co-expression analysis was determined by Spearman correlation coefficients and displayed in the form of volcano and heat maps. Gene ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of co-expressed genes were performed by cluster profile package of R software, and we performed a visualization analysis of the data using the ggplot2 package. The PPI interaction network was constructed by Search Tool for the Retrieval of Interacting Genes (STRING) database (http://string-db.org/).

Gene set enrichment analysis (GSEA)

GSEA is an approach that focuses on gene sets to explain biological pathways enriched by different populations (Subramanian et al., 2005). The samples were divided into high and low expression groups according to the median LECT2 expression, and then GSEA functional analysis was performed using the "limma", "enrichplot", "clusterProfiler" and "org.Hs.eg.db" packages. Gene set "c2. cp.kegg.v7.4. symbols.gmt" are obtained from GSEA website (https://www.gsea-msigdb.org/gsea/index.jsp).

Immune infiltrate analysis

For a more comprehensive estimation of immune cell infiltration, we applied seven algorithms to estimate the immune cell infiltration status in the samples, including XCELL, TIMER, QUANTISEQ, MCPCOUNTER, EPIC, CIBERSORT abs, and CIBERSORT. These immune cell infiltration level results can be obtained from TIMER2.0 (Li et al., 2020) (http://timer.cistrome.org/). The relationship between LECT2 expression and immune infiltrating cells was calculated using Spearman correlation analysis. The significance threshold was set at $p < 0.05$.

Correlation between LECT2 expression with immune checkpoint molecules and HLA genes

We calculated the relationship between LECT2 expression with 48 immune checkpoint molecules and 19 HLA genes using Spearman correlation analysis.

Chemotherapy drug sensitivity analysis

We calculated the semi-inhibitory concentrations (IC50) of six commonly used chemotherapeutic drugs using the pRRophetic package to evaluate the sensitivity of HCC samples to the six chemotherapeutic drugs. IC50 difference between low and high expression groups was compared using Wilcoxon signed-rank test.

Immunohistochemistry (IHC)

We examined 72 paraffin-embedded HCC tissues and adjacent tissue samples from the First Affiliated Hospital of Anhui Medical University using immunohistochemistry. These tissue specimens were obtained from patients who underwent liver resection from 2014 to 2015. All patients provided written informed consent and adhered to the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the First Affiliated Hospital of Anhui Medical University. Two experienced pathologists independently calculated immunohistochemical scores A score greater than or equal to three was considered high expression and less than three was considered low expression.

Statistical analysis

In the comparison of clinicopathological features. Student's t-test and Chi-square test were performed according to different types of variables. The results were considered statistically significant at a two-sided $p < 0.05$.

Result

LECT2 expression levels comparison

The full names and abbreviations of the 33 cancers are shown in Supplementary Table S1. We analyzed the expression levels of LECT2 in 33 cancers in the TCGA database. Compared with the normal tissues, the results showed that LECT2 was differentially expressed in 14 of the 33 cancers (BLCA、BRCA、CHOL、ESCA、GBM、HNSC、KICH、KIRC、HCC、LUAD、LUSC、PCPG、PRAD、UCES). In particular, the expression levels of LECT2 were decreased in HCC and CHOL (Figure 1A). Meanwhile, we ranked the expression levels of LECT2 in 33 cancers and found that LECT2 had higher expression levels only in HCC and CHOL (Figure 1B).

Prognostic value of LECT2

We further analyzed the prognostic value of LECT2 in cancers with differential expression. Because LECT2 is expressed only at high levels in CHOL and HCC and is too low in other cancers, we focused on the prognostic value of LECT2 in CHOL and HCC. Patients were divided into high

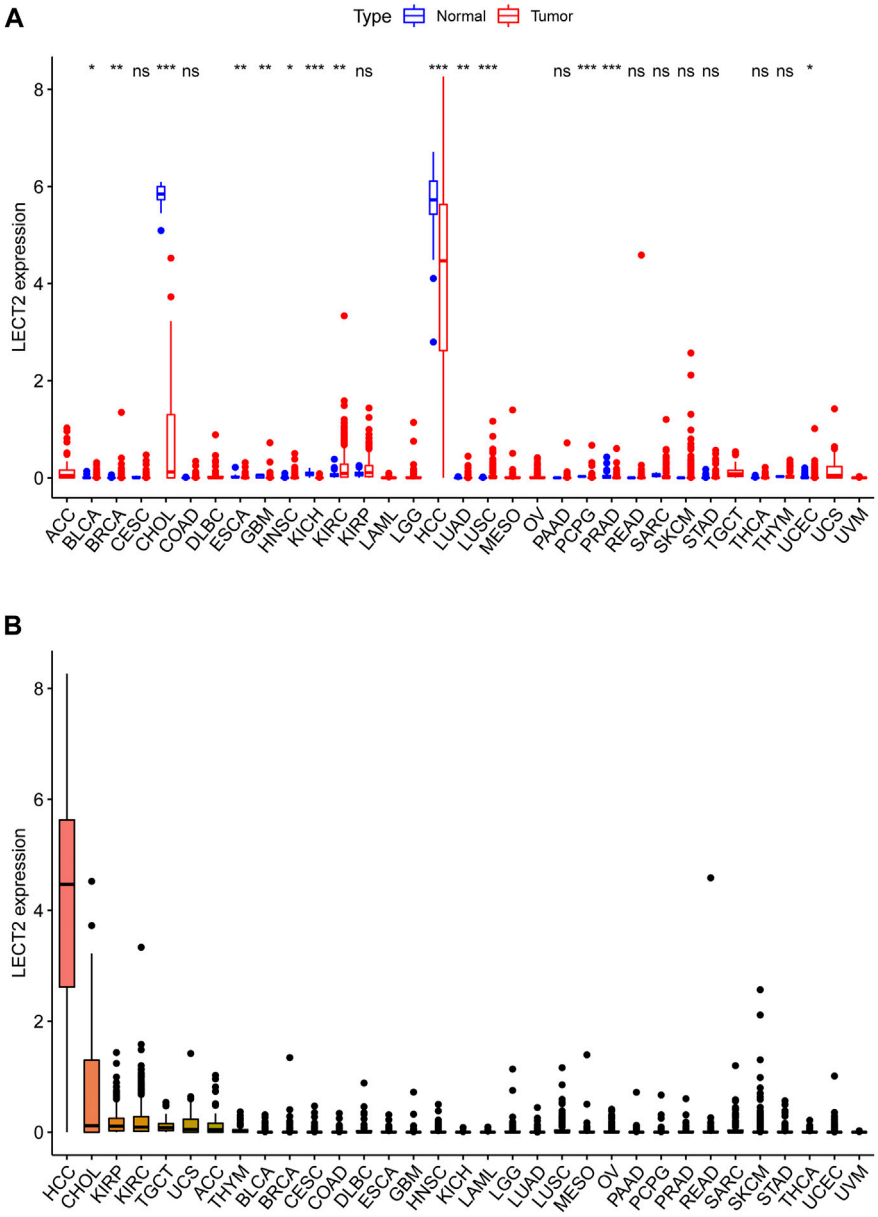


FIGURE 1
LECT2 expression levels comparison. **(A)** The differential expression analysis between tumor and normal groups of LECT2 in 33 cancers **(B)** The expression of LECT2 in 33 cancers (from high to low). "ns" represents no significance, "*" represents $p < 0.05$, "**" represents $p < 0.01$, and "***" represents $p < 0.001$ ".

expression group and low expression group, and we used Kaplan-Meier survival analysis to draw the survival curve of patients. In addition to OS, we also explored other important prognostic indicators such as DFS, DSS, and PFS. The results showed that the expression level of LECT2 did not affect the prognostic profile of CHOL patients, including OS ($p = 0.106$, Figure 2A), DFS ($p = 0.236$, Figure 2B), DSS ($p = 0.138$, Figure 2C) and PFS ($p = 0.147$, Figure 2D). However, the expression level of LECT2 significantly affected the prognosis of HCC patients. Patients with high expression of LECT2 had better OS ($p < 0.001$, Figure 2E), DFS ($p < 0.001$, Figure 2F), DSS ($p = 0.004$, Figure 2G) and PFS ($p = 0.002$, Figure 2H). The results showed that LECT2 is associated with the prognosis of HCC.

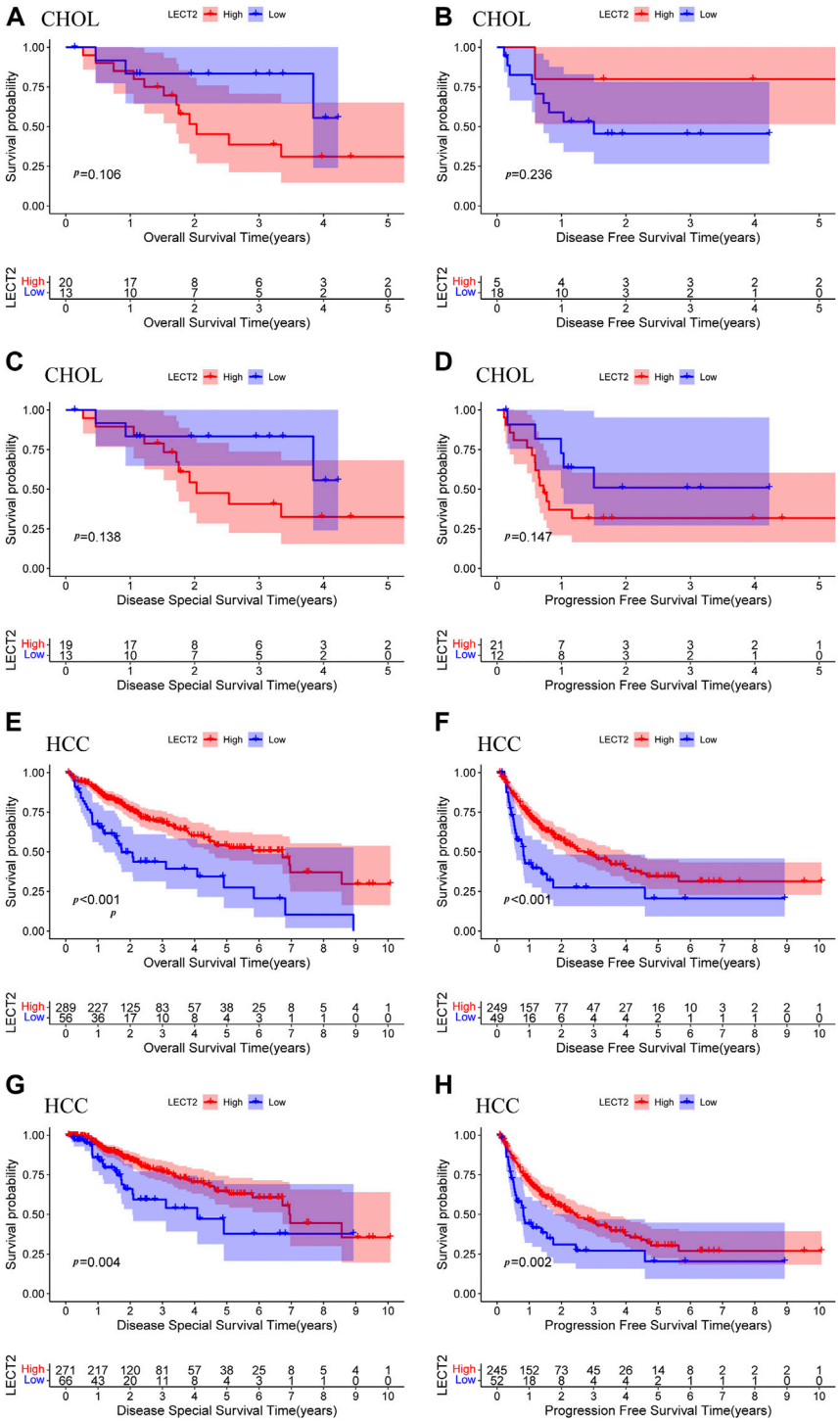


FIGURE 2
Prognostic analysis of LECT2. OS **(A)**, DFS **(B)**, DSS **(C)**, and PFS **(D)** of CHOL patients were grouped by high and low expression of LECT2. OS **(E)**, DFS **(F)**, DSS **(G)**, and PFS **(H)** of HCC patients were grouped by high and low expression of LECT2. OS: overall survival, DFS: disease-free survival, DSS: disease-specific survival, PFS: progression-free survival.

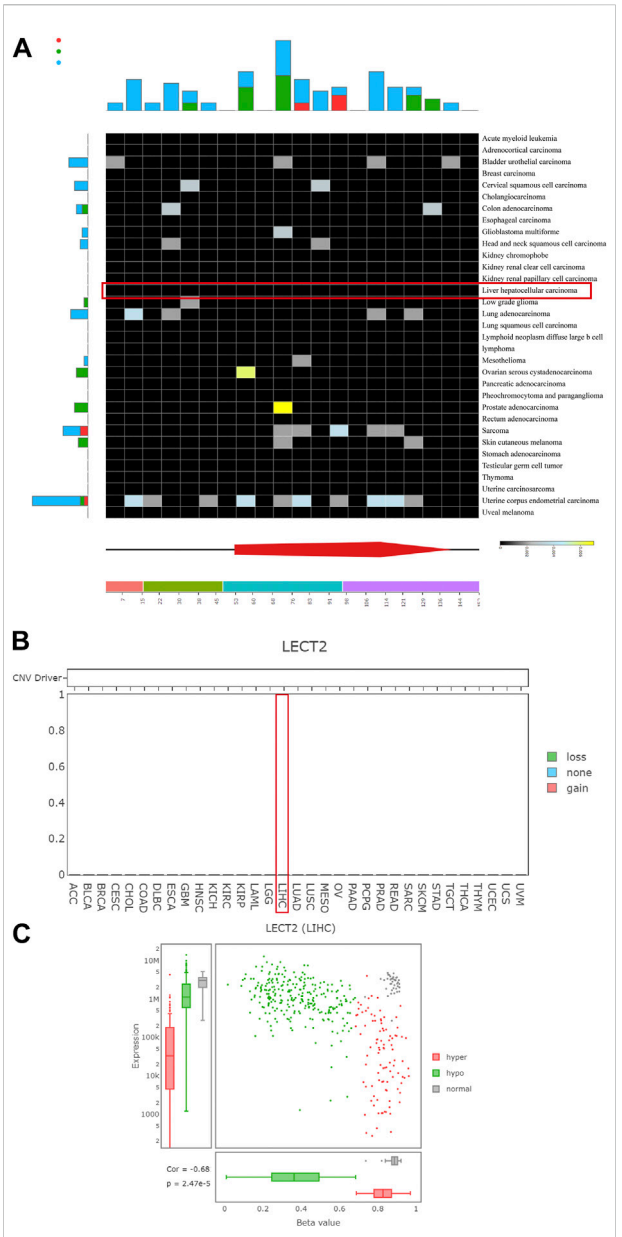


FIGURE 3
Mutation, CNV, and methylation profiles of LECT2. **(A)** The mutation status of LECT2 in various tumors according to the DriverDBv3 database **(B)** The CNV of LECT2 in various tumors **(C)** The methylation of LECT2 in HCC.

Mutation, CNV, and methylation profiles of LECT2

We further analyzed the possible reasons for the differential expression of LECT2 in HCC. To comprehensively analyze the mutational, CNV, and methylation spectrum of LECT2, we used DriverDBv3 to explore the mutational and CNV in all cancer types in the TCGA database. The DriverDBv3 database results showed no mutations and CNV of LECT2 in HCC (Figures 3A, B). Then we explored the correlation between LECT2 expression levels and methylation levels in HCC. The results showed that methylation levels were negatively correlated with LECT2 expression, with hypermethylation usually implying lower LECT2 expression ($p < 0.001$, Figure 3C). We think that methylation is one of the possible reasons for differential expression of LECT2 in HCC.

Expression, prognostic value and clinicopathological features of LECT2 in HCC

Given the unique expression pattern and prognostic value of LECT2 in HCC, we further validated the expression and prognostic value of LECT2 in HCC. representative IHC maps of LECT2 in HCC tissues and adjacent normal tissues are shown in Figures 4A,B. Further we examined the expression levels of LECT2 in 72 pairs of HCC and adjacent tissues. Consistent with the results of TCGA database, we found that the expression level of LECT2 was higher in adjacent tissues than in HCC tissues (Figure 4C). Meanwhile, we analyzed the prognostic and clinicopathological characteristics of 72 samples. The results showed that HCC patients in the high LECT2 expression group had better OS ($p = 0.003$, Figure 4D). In addition, the results showed that LECT2 expression levels correlated with TNM stage ($p = 0.011$), and TNM stage was higher in the low LECT2 expression group (Table 1).

Enrichment analysis of LECT2 in HCC and PPI network analysis

To find out the biological role of LECT2 in HCC, we analyzed the co-expression of LECT2 in HCC using the LinkedOmics database. As shown in Figure 5A, 3348 genes are positively related to LECT2, and 7,873 genes are negatively related to LECT2 ($p < 0.05$). The first 50 important genes that are positively (Figure 5B) and negatively (Figure 5C) correlated with LECT2 are shown in the heat map. After the GO and KEGG analysis of the top 200 co-expressed genes positively related to LECT2 expression. The top 15 results for GO including Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) are shown in the bubble chart. Similarly, it showed the top 10 results for KEGG. The results of GO function showed that LECT2 co-expression is enriched in the small molecule catabolic process and cellular amino acid metabolic process (Figure 5D). The results of KEGG showed that LECT2 co-expression is correlated with Fatty acid degradation, Pyruvate metabolism, PPAR signaling pathway, and Peroxisome (Figure 5E). We studied the PPI network of LECT2 using the STRING database to learn more about the potential mechanisms of action of LECT2. We found that

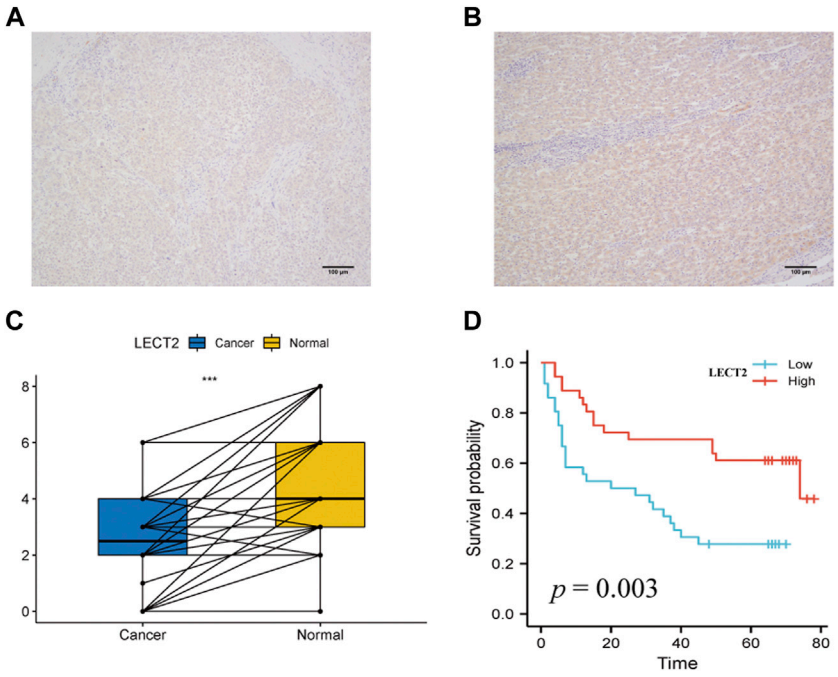


FIGURE 4
Expression levels and prognostic analysis of LECT2. **(A–B)** Representative IHC maps of LECT2 expression in HCC tissues and adjacent normal tissues **(C)** The expression levels of LECT2 in 72 cases of HCC and normal tissues, blue represents HCC tissue, yellow represents adjacent normal tissues **(D)** The prognostic analysis of LECT2 in 72 cases of HCC.

TABLE 1 Clinicopathological features of LECT2 in HCC.

Factor	IHC score of LECT2		*p*-value
	High (n = 36)	Low (n = 36)	
Age, year	56.33 ± 11.50	57.67 ± 12.80	0.644
Sex			
Male	30 (83.3%)	30 (83.3%)	1.000
Female	6 (16.7%)	6 (16.7%)	
TNM stage			
I + II	33 (91.7%)	23 (63.9%)	**0.011**
III + IV	3 (8.3%)	13 (36.1%)	
Grade			
G1	8 (22.2%)	5 (13.9%)	0.634
G2	19 (52.8%)	22 (61.1%)	
G3	9 (25.0%)	9 (25.0%)	
The longest diameter of tumor	4.97 ± 3.35	5.89 ± 3.33	0.251

TNM, Tumor-Node-Metastasis. IHC, Immunohistochemistry.

LECT2 is mainly related to LECT1 (0.775), SVIL (0.642), FGG (0.610), and DDX46 (0.601) which are the first four proteins (Figure 5F). The results showed that the expression of LECT2 is associated with several metabolic and disease pathways.

GSEA of LECT2

Considering the strong correlation between LECT2 and HCC, we decided to investigate the potential pathways of

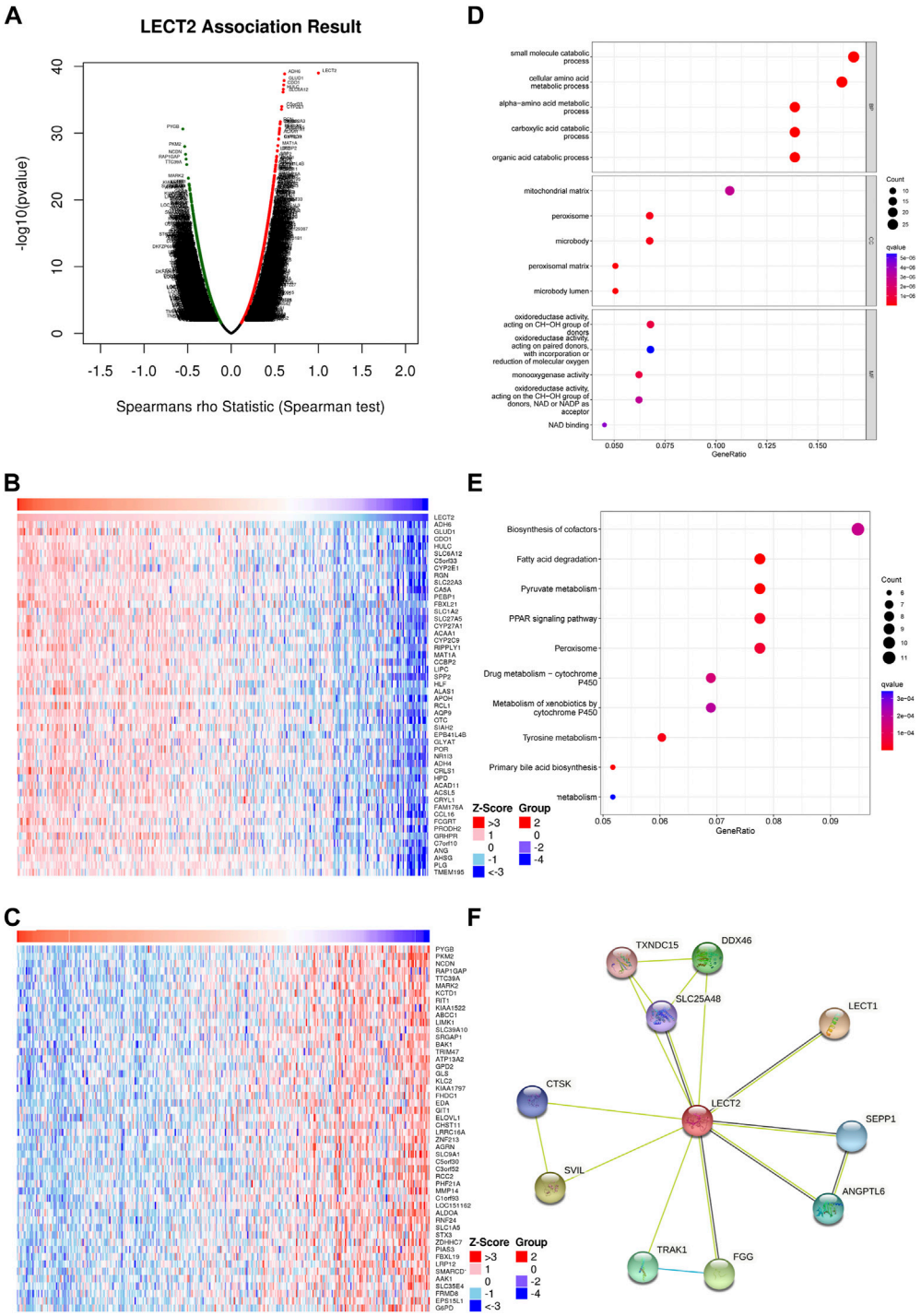


FIGURE 5
Enrichment analysis and PPI analysis of LECT2 in HCC. **(A)** Genes highly correlated with LECT2 identified in HCC by Spearman correlation analysis **(B)** Top 50 genes positively correlated with LECT2 in HCC. **(C)** Top 50 genes negatively correlated with LECT2 in HCC **(D)** Enrichment of gene ontology (GO) for genes correlated with LECT2. **(E)** Enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) for genes correlated with LECT2 **(F)** Protein-protein interaction network of LECT2.

LECT2 dysregulation in HCC.HCC patients were divided into high- and low-expression groups according to the median mRNA expression of LECT2 in the HCC cohort in TCGA. Further functional enrichment analyses showed in the LECT2 high expression group, the five most significant pathways, including KEGG DRUG METABOLISM

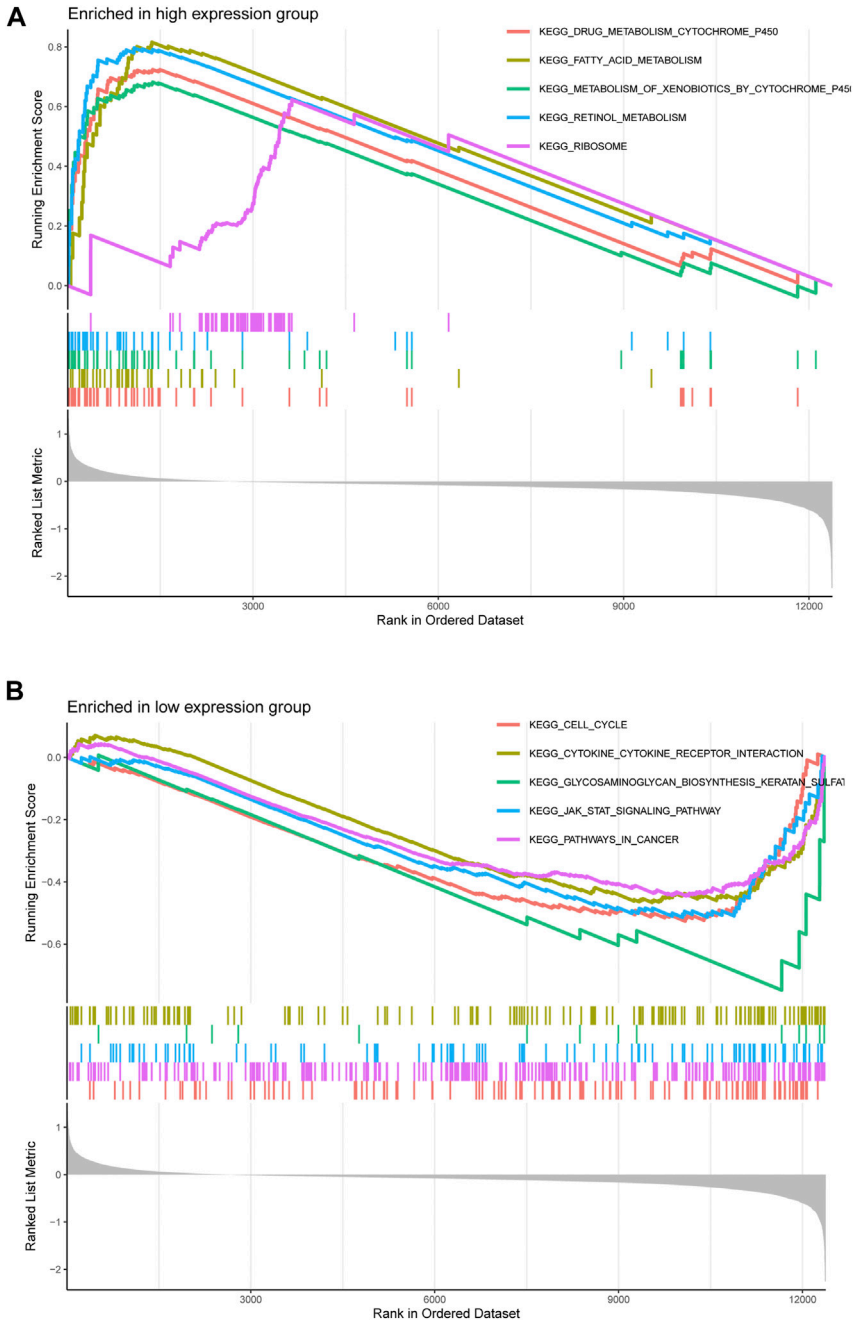


FIGURE 6
Gene Set Enrichment Analysis of LECT2 in HCC. **(A)** The five pathways were most significantly enriched in the LECT2 high-expression group **(B)** The five pathways were most significantly enriched in the LECT2 low-expression group.

CYTOCHROME P450, KEGG FATTY ACID METABOLISM, KEGG METABOLISM OF XENOBIOTICS BY CYTOCHROME P450, KEGG RETINOL METABOLISM and KEGG RIBOSOME were enriched (Figure 6A). In the LECT2 low expression group, the five most significant pathways, including KEGG CELL CYCLE, KEGG CYTOKINE RECEPTOR INTERACTION, KEGG GLYCOSAMINOGLYCAN BIOSYNTHESIS KERATAN SULFAT, KEGG JAK STAT SIGNALING PATHWAY and KEGG PATHWAYS IN CANCER were enriched (Figure 6B).

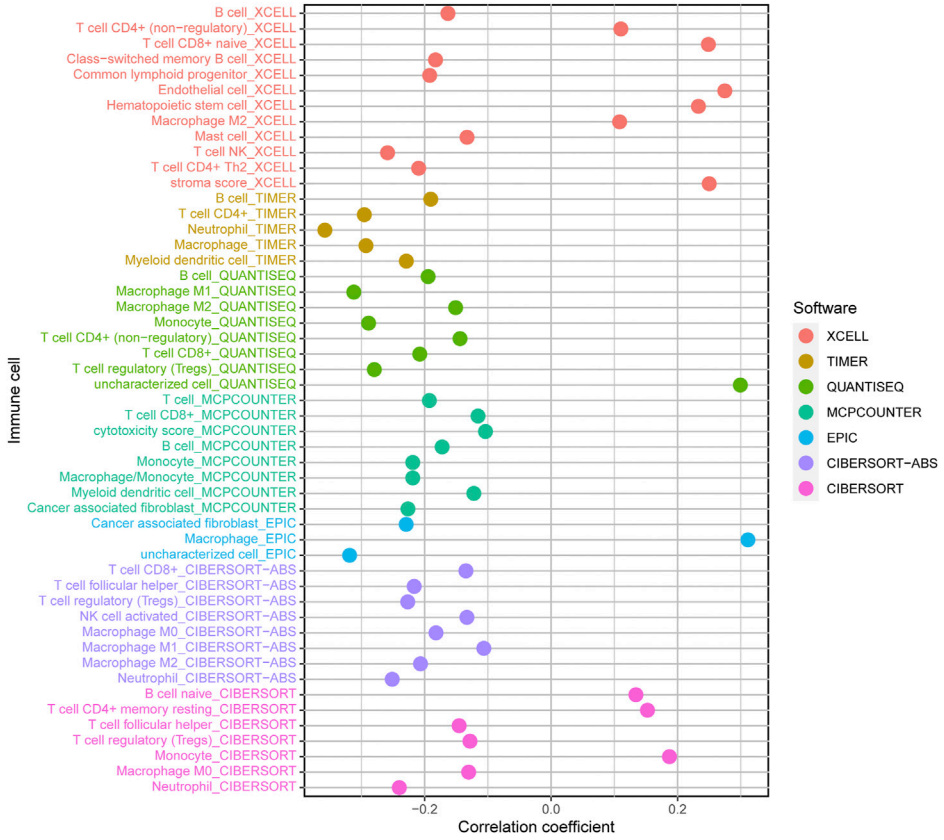


FIGURE 7
Correlations of LECT2 expression with immune infiltration level in HCC.

Detailed GSEA analysis information is shown in Supplementary Table S2.

Correlation between LECT2 and immune cells infiltrating

We further explored the relationship between LECT2 and the tumor immune microenvironment. The results showed that LECT2 expression was negatively correlated with immune infiltration of B cells, Neutrophil, Monocyte, Cancer-associated fibroblast, and Myeloid dendritic cell, and positively correlated with T cell $CD8^+$ naive, Endothelial cell, and Hematopoietic stem cell. Meanwhile, the relationship between LECT2 expression and macrophages and macrophage M2 showed different results in different methods (Figure 7). The results suggested that Lect2 expression may affect the level of multiple immune cell infiltration in the tumor microenvironment of HCC.

Correlation between LECT2 and predictive immune markers (Checkpoint and HLA) molecules

In addition, immunotherapy targeting immune checkpoint molecules is a promising target for immunotherapy in HCC patients. We then analyzed the relationship between LECT2 expression and 48 immune checkpoint molecules. We found that LECT2 was positively correlated with two immune checkpoint molecules, and negatively correlated with 31 immune checkpoint molecules (Figure 8). Therefore, HCC patients in the LECT2 low expression group may be more sensitive to immune checkpoint inhibitors, such as PD1 inhibitors and CTLA-4 inhibitors. HLA genes are important immune genes in the human body, and tumor-induced immune escape can change the expression of the HLA gene so that the tumor can evade the immune system without being killed (McGranahan et al., 2017). The results showed that LECT2 was negatively correlated with 18 of 19 HLA genes (Figure 9). In summary, LECT2 was negatively correlated with most of immune checkpoint molecules and HLA genes.

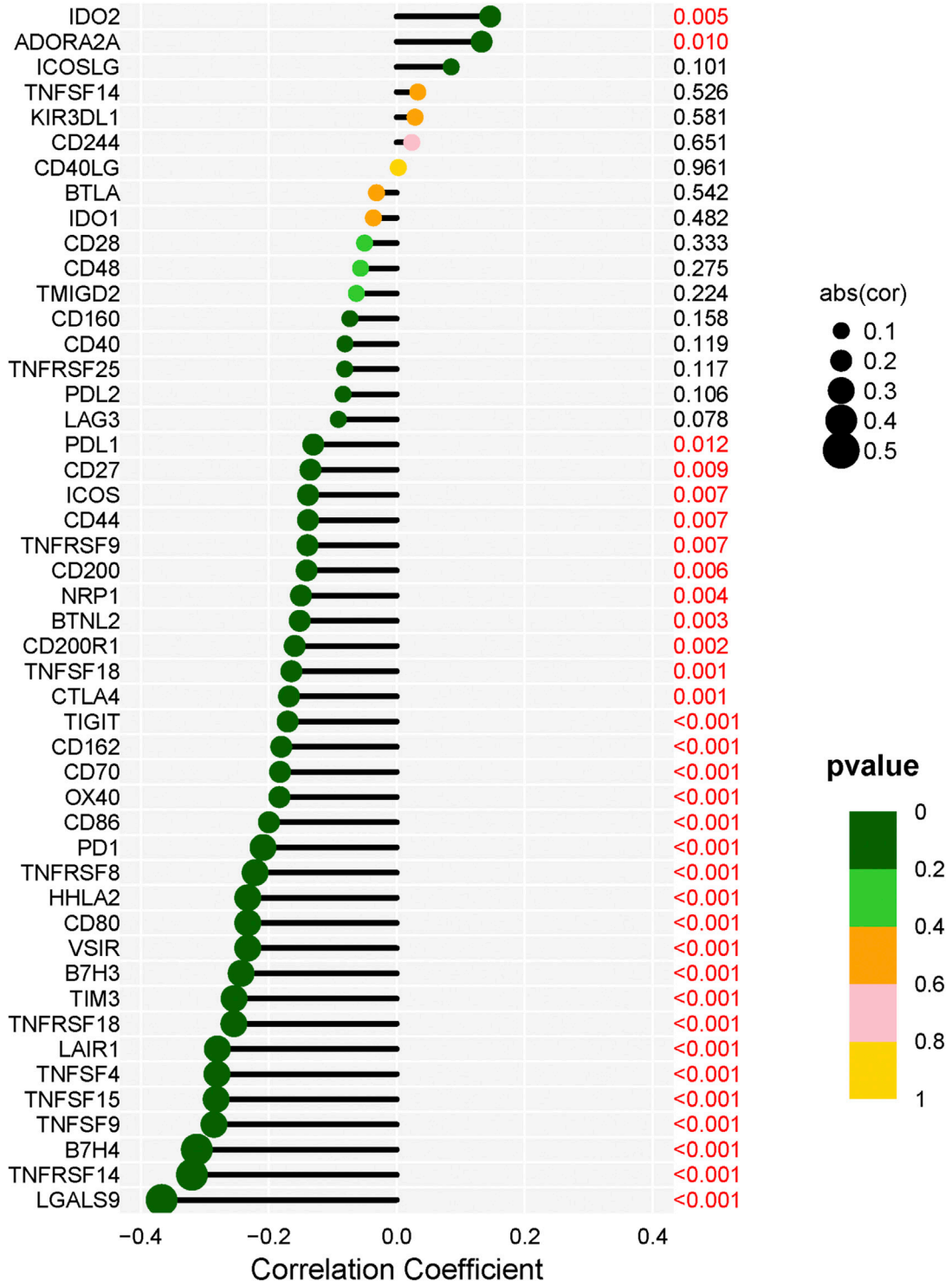


FIGURE 8
Correlation analysis of LECT2 expression with 48 immune checkpoint molecules.

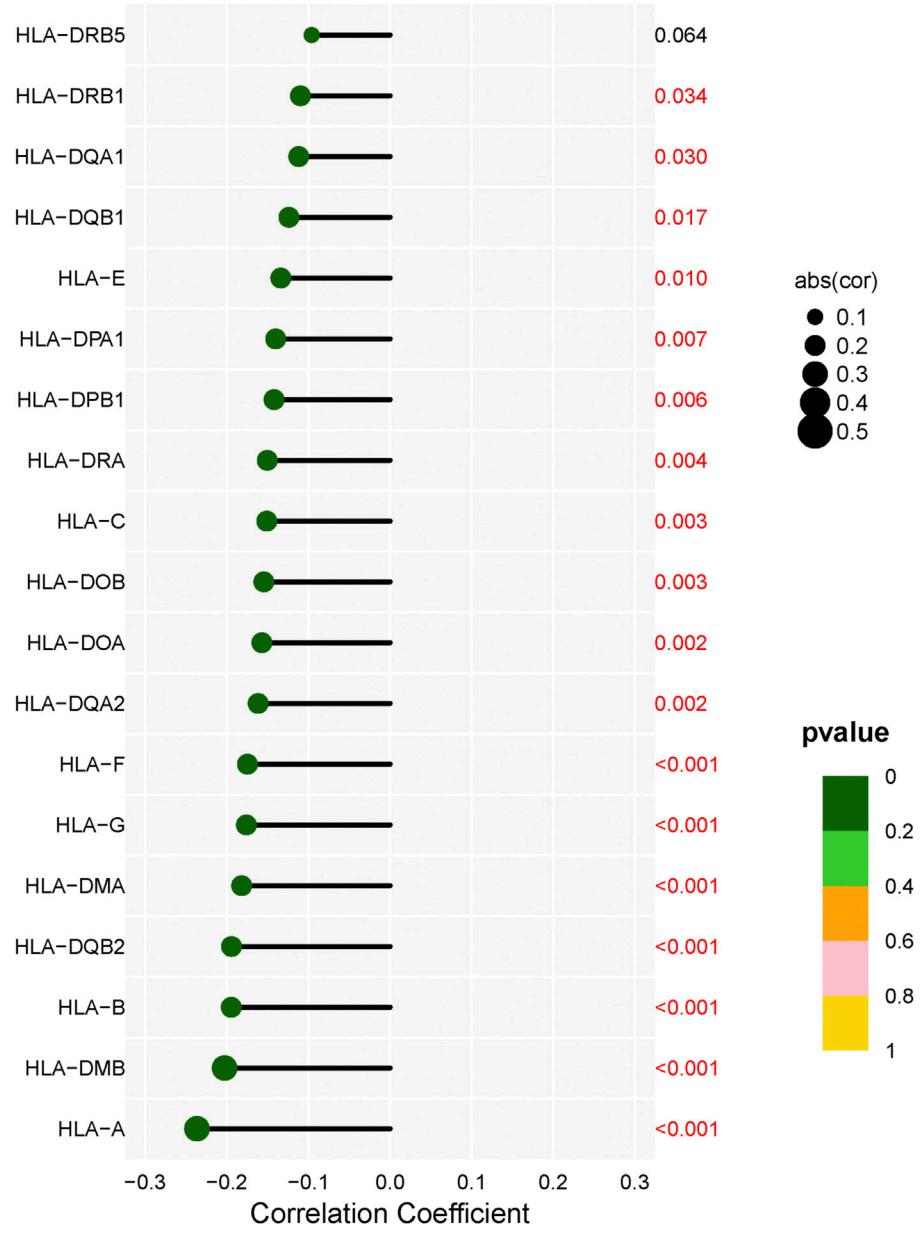


FIGURE 9
Correlation analysis of LECT2 expression with 19 HLA genes.

Correlation analysis of LECT2 expression with chemotherapy drugs

In addition, we also explored the relationship between the expression of LECT2 and the sensitivity of HCC patients to several common chemotherapeutic drugs. Sorafenib (Figure 10A), Cisplatin (Figure 10B), Rapamycin (Figure10C), Mitomycin (Figure 10D), Doxorubicin (Figure 10E), Bleomycin (Figure 10F). The results showed that high expression of LECT2 was associated with higher IC50 of Cisplatin ($p < 0.01$), Rapamycin ($p < 0.001$), and Mitomycin. C ($p < 0.05$) chemotherapy drugs (Figure 10). These results implied that patients with different LECT2 expression levels have different sensitivities to a variety of common chemotherapeutic drugs.

Discussion

LECT2 is a 16-kDa secreted protein. It is mainly produced by hepatocytes (Yamagoe et al., 1998) and is usually expressed in

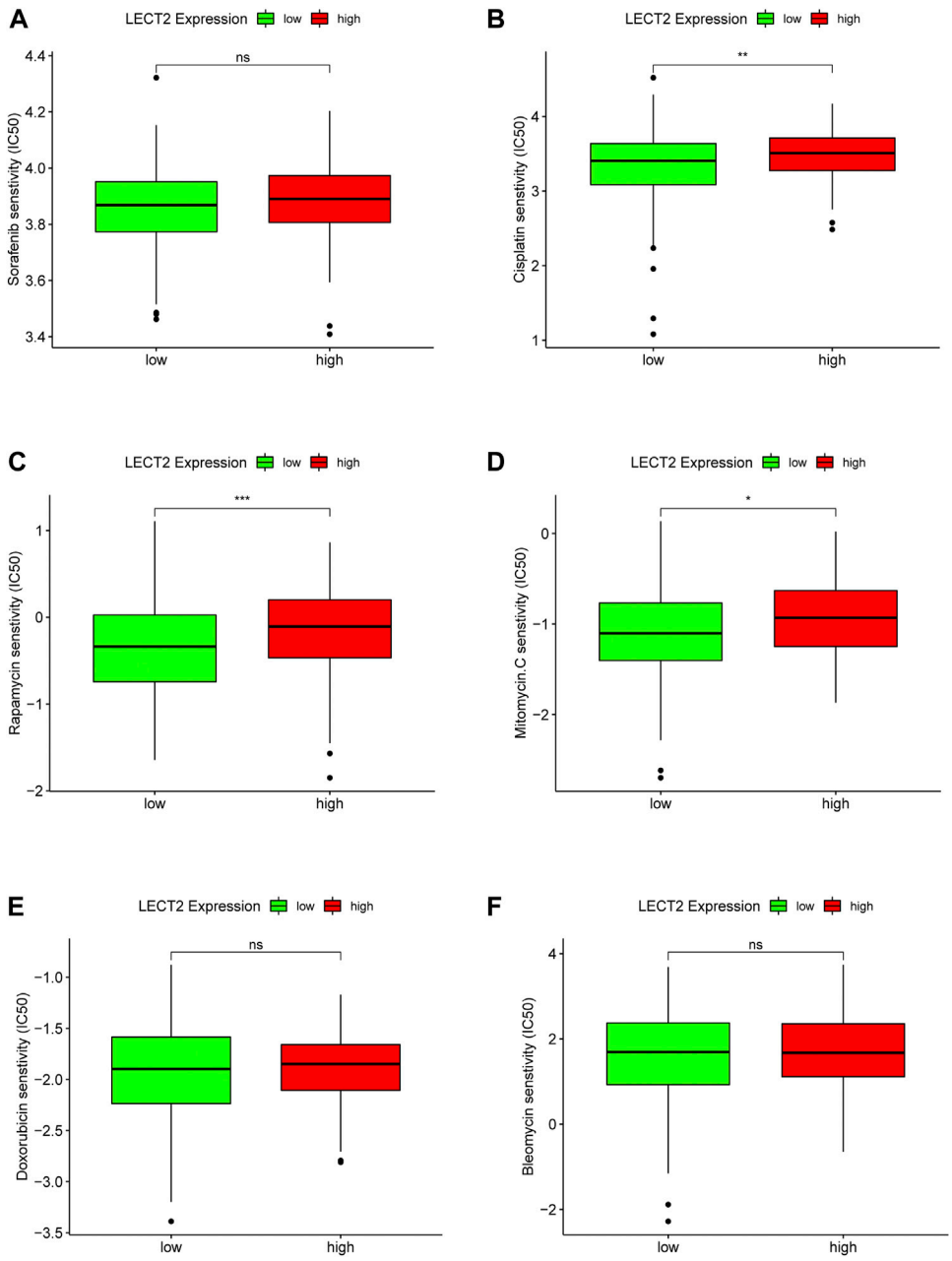


FIGURE 10
Correlation analysis of LECT2 expression with chemotherapeutic drug sensitivity. Difference analysis of the sensitivity of six chemotherapeutic drugs **(A)** Sorafenib, **(B)** Cisplatin, **(C)** Rapamycin, **(D)** Mitomycin, **(E)** Doxorubicin, **(F)** Bleomycin in LECT2 high expression group and low expression group, green represents low LECT2 expression, red represents high LECT2 expression. "ns" represents no significance, "*" represents $p < 0.05$, "**" represents $p < 0.01$, and "***" represents $p < 0.001$ ".

vascular cells, endothelial cells, and VSMC (Slowik and Apte, 2017). A large number of studies have now shown that LECT2 is associated with the progression of a variety of cancers. For example, LECT2 is considered one of the potential prognostic risk biomarkers for colon adenocarcinoma (Yin et al., 2020). By inhibiting angiogenesis, LECT2 inhibits tumor growth in HCC (Chen et al., 2016). HCC with low LECT2 expression has a higher grade and inflammatory infiltrates (L'Hermitte et al., 2019). However, there are no systematic and comprehensive studies on the role of LECT2 in HCC. Therefore, there is a need to further explore the potential mechanisms of LECT2 in HCC.

According to our knowledge, this is the first study to assess the role and significance of LECT2 in HCC regarding clinical, biological, and genomic aspects, laying the foundation for the clinical application of LECT2.

In the present study, we found higher expression levels of LECT2 in HCC and CHOL, and we reported significantly lower expression levels of LECT2 in HCC and CHOL samples compared to normal tissue. We further analyzed the prognostic value of LECT2 in HCC and CHOL. We found that LECT2 could affect both OS, DFS, DSS, and PFS in HCC patients and that low expression of LECT2 levels was a factor in poorer prognosis in HCC patients. Furthermore, we found that LECT2 mutations and CNV are uncommon times in HCC, but the abnormal expression of LECT2 may be due to abnormal methylation. Then, the results of 72 HCC clinical samples were consistent with the TCGA database, with HCC tissues having lower LECT2 levels than adjacent tissues. And the results showed that HCC patients in the LECT2 high expression group had better OS. Analysis of clinicopathological features showed a lower TNM stage in the group with high expression of LECT2. This suggests that high levels of LECT2 inhibit the progression of HCC.

Lu et al. found that LECT2 may inhibit HCC cell glycolysis during aerobic glycolysis, and reduced glycolysis by LECT2 might be linked to the inhibitory effect on HCC cells (Lu et al., 2020). However, there have been no studies on the functional enrichment analysis of LECT2 co-expression in HCC. In the present study, by GO and KEGG analysis of 200 genes associated with LECT2, we found that co-expression of LECT2 was mainly enriched in the small molecule catabolic process and cellular amino acid metabolic process. And we found that the LECT2 co-expression was mainly related to the Fatty acid degradation, Pyruvate metabolism, PPAR signaling pathway, and Peroxisome by KEGG analysis. It has been shown that the PPAR signaling pathway plays a key role in tumors (Wagner and Wagner, 2020). PPI analysis found that LECT2 has the strongest correlation with LECT1, SVIL, FGG, and DDX46. The study of LECT1 on osteosarcoma cells *in vivo* showed that it inhibited their growth and proliferation (Lin et al., 2017). Knockdown of DDX46 inhibited osteosarcoma cell proliferation and tumor growth *in vivo* (Jiang et al., 2017). SVIL (Houlier et al., 2020) and FGG (Peng et al., 2021) also have corresponding roles in the tumor process. At the same time, the GSEA pathway enrichment analysis showed that the JAK/STAT signaling pathway, cell cycle, and pathways in cancer were enriched in the low LECT2 expression group. Interestingly, blockade of the JAK/STAT signaling pathway mediated by SOCS3 was recently reported to inhibit the progression of HCC (Liu et al., 2021). These pathways may be potential mechanisms for LECT2 to regulate HCC. We suggest that LECT2 has other biological functions in HCC besides participating in glycolysis in HCC.

It is well documented that the tumor microenvironment plays an indispensable role in malignant tumors, and among them, immune cells are significant, and the level of various tumor immune cells affects the therapeutic effect (Hinshaw and Shevde, 2019). Therefore, the present study focused on exploring the correlation between tumor immune infiltration and LECT2. We used seven common methods for assessing immune cell infiltration. The results showed that LECT2 expression was negatively correlated with immune infiltration of B cells, Neutrophil, Monocyte, Cancer-associated fibroblast, and Myeloid dendritic cell, and positively correlated with T cell $CD8^+$ naive, Endothelial cell, and Hematopoietic stem cell. The cancer-associated fibroblasts can increase angiogenesis, inflammation, proliferation, survival, EMT, and alter immune surveillance to promote HCC (Affo et al., 2017). Similarly, LECT2 loss contributes to the proliferation of inflammatory monocytes in HCC (L'Hermitte et al., 2019). These results are consistent with our analysis of the tumor suppressive role played by LECT2 in HCC, suggesting that LECT2 may regulate the progression of HCC by affecting these immune cells. In addition, our results showed that LECT2 was negatively correlated with 31 immune checkpoint molecules, including PD1 and CTLA-4, and was negatively correlated with 18 of 19 HLA genes. Moreover, in the LECT2 high expression group, the IC50 of chemotherapy drugs such as Cisplatin, Rapamycin, and Mitomycin. C was increased. In conclusion, these results provide a reference for the clinical use of drugs in HCC patients.

In conclusion, by comprehensively elucidating the expression, prognostic value, association with clinicopathological factors, co-expression network, pathway enrichment analysis, and crosstalk with immune infiltration in HCC, LECT2 may be a new potential prognostic and diagnostic biomarker for hepatocellular carcinoma with potential clinical applications.

Conclusion

This study is the first to provide a comprehensive and detailed analysis of the role of LECT2 in HCC and to show that LECT2 is a new potential diagnostic and prognostic biomarker for hepatocellular carcinoma. However, further research is needed to explain the mechanisms of LECT2 involvement in HCC.

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.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical approval was obtained from the Ethics Committee of the First Affiliated Hospital of Anhui Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

H-HX and Y-FG contributed to the study concept and design. J-FQ, W-JS, H-Z, W-H W, and Z-HW contributed to the acquisition of data, and W-JS, J-FQ, and J-PS contributed to statistical analysis and data interpretation. J-FQ, Y-Y W, and H-Z wrote the first draft of the manuscript. H-HX, Y-FG, and Y-YL supervised and oversaw the study. All authors contributed to manuscript revision, read and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (Youth Program) (81900554), the Major Program for Supporting Outstanding Talents in Colleges of the Ministry of Human Resources and Social Security of the People's Republic of Anhui (gxyqZD2020013), and the National Natural Science Foundation of Anhui Province (2208085MH204).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.951077/full#supplementary-material

References

Affo, S., Yu, L. X., and Schwabe, R. F. (2017). The role of cancer-associated fibroblasts and fibrosis in liver cancer. *Annu. Rev. Pathol.* 12, 153–186. doi:10.1146/annurev-pathol-052016-100322

Ando, K., Kato, H., Kotani, T., Ozaki, M., Arimura, Y., and Yagi, J. (2012). Plasma leukocyte cell-derived chemotaxin 2 is associated with the severity of systemic inflammation in patients with sepsis. *Microbiol. Immunol.* 56 (10), 708–718. doi:10.1111/j.1348-0421.2012.00488.x

Bengtsson, A., Andersson, R., and Ansari, D. (2020). The actual 5-year survivors of pancreatic ductal adenocarcinoma based on real-world data. *Sci. Rep.* 10 (1), 16425. doi:10.1038/s41598-020-73525-y

Chen, C. K., Yu, W. H., Cheng, T. Y., Chen, M. W., Su, C. Y., Yang, Y. C., et al. (2016). Inhibition of VEGF165/VEGFR2-dependent signaling by LECT2 suppresses hepatocellular carcinoma angiogenesis. *Sci. Rep.* 6, 31398. doi:10.1038/srep31398

Comenzo, R. L. (2014). LECT2 makes the amyloid list. *Blood* 123 (10), 1436–1437. doi:10.1182/blood-2014-01-549758

Greenow, K. R., Zverev, M., May, S., Kendrick, H., Williams, G. T., Phesse, T., et al. (2018). Lect2 deficiency is characterised by altered cytokine levels and promotion of intestinal tumourigenesis. *Oncotarget* 9 (92), 36430–36443. doi:10.18632/oncotarget.26335

Hinshaw, D. C., and Shevde, L. A. (2019). The tumor microenvironment innately modulates cancer progression. *Cancer Res.* 79 (18), 4557–4566. doi:10.1158/0008-5472.CAN-18-3962

Houlier, A., Pissaloux, D., Masse, I., Tirode, F., Karanian, M., Pincus, L. B., et al. (2020). Melanocytic tumors with MAP3K8 fusions: Report of 33 cases with morphological-genetic correlations. *Mod. Pathol.* 33 (5), 846–857. doi:10.1038/s41379-019-0384-8

Ito, M., Nagata, K., Kato, Y., Oda, Y., Yamagoe, S., Suzuki, K., et al. (2003). Expression, oxidative refolding, and characterization of six-histidine-tagged recombinant human LECT2, a 16-kDa chemotactic protein with three disulfide bonds. *Protein Expr. Purif.* 27 (2), 272–278. doi:10.1016/s1046-5928(02)00634-4

Jiang, F., Zhang, D., Li, G., and Wang, X. (2017). Knockdown of DDX46 inhibits the invasion and tumorigenesis in osteosarcoma cells. *Oncol. Res.* 25 (3), 417–425. doi:10.3727/096504016X14747253292210

L'Hermitte, A., Pham, S., Cadoux, M., Couchy, G., Caruso, S., Anson, M., et al. (2019). Lect2 controls inflammatory monocytes to constrain the growth and progression of hepatocellular carcinoma. *Hepatology* 69 (1), 160–178. doi:10.1002/hep.30140

Lan, F., Misu, H., Chikamoto, K., Takayama, H., Kikuchi, A., Mohri, K., et al. (2014). LECT2 functions as a hepatokine that links obesity to skeletal muscle insulin resistance. *Diabetes* 63 (5), 1649–1664. doi:10.2337/db13-0728

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 48 (W1), W509–W514. doi:10.1093/nar/gkaa407

Lin, X., Wang, L., and Wang, F. (2017). ChondromodulinI suppresses tumorigenesis of human osteosarcoma cells. *Mol. Med. Rep.* 16 (6), 8542–8548. doi:10.3892/mmr.2017.7629

Liu, Z. K., Li, C., Zhang, R. Y., Wei, D., Shang, Y. K., Yong, Y. L., et al. (2021). EYA2 suppresses the progression of hepatocellular carcinoma via SOCS3-mediated blockade of JAK/STAT signaling. *Mol. Cancer* 20 (1), 79. doi:10.1186/s12943-021-01377-9

Lu, C., Fang, S., Weng, Q., Lv, X., Meng, M., Zhu, J., et al. (2020). Integrated analysis reveals critical glycolytic regulators in hepatocellular carcinoma. *Cell. Commun. Signal.* 18 (1), 97. doi:10.1186/s12964-020-00539-4

McGranahan, N., Rosenthal, R., Hiley, C. T., Rowan, A. J., Watkins, T. B. K., Wilson, G. A., et al. (2017). Allele-specific HLA loss and immune escape in Lung cancer evolution. *Cell.* 171 (6), 1259–1271. e1211. doi:10.1016/j.cell.2017.10.001

Peng, H. H., Wang, J. N., Xiao, L. F., Yan, M., Chen, S. P., Wang, L., et al. (2021). Elevated serum FGG levels prognosticate and promote the disease progression in prostate cancer. *Front. Genet.* 12, 651647. doi:10.3389/fgene.2021.651647

Slowik, V., and Apte, U. (2017). Leukocyte cell-derived chemotaxin-2: It's role in pathophysiology and future in clinical medicine. *Clin. Transl. Sci.* 10 (4), 249–259. doi:10.1111/cts.12469

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Takata, N., Ishii, K. A., Takayama, H., Nagashimada, M., Kamoshita, K., Tanaka, T., et al. (2021). LECT2 as a hepatokine links liver steatosis to inflammation via activating tissue macrophages in NASH. *Sci. Rep.* 11 (1), 555. doi:10.1038/s41598-020-80689-0

Vasaikar, S. V., Straub, P., Wang, J., and Zhang, B. (2018). LinkedOmics: Analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* 46 (D1), D956–D963. doi:10.1093/nar/gkx1090

Wagner, N., and Wagner, K. D. (2020). PPAR beta/delta and the hallmarks of cancer. *Cells* 9 (5), E1133. doi:10.3390/cells9051133

Xu, M., Xu, H. H., Lin, Y., Sun, X., Wang, L. J., Fang, Z. P., et al. (2019). LECT2, a ligand for Tie1, plays a crucial role in liver fibrogenesis. *Cell.* 178 (6), 1478–1492. doi:10.1016/j.cell.2019.07.021

Yamagoe, S., Mizuno, S., and Suzuki, K. (1998). Molecular cloning of human and bovine LECT2 having a neutrophil chemotactic activity and its specific expression in the liver. *Biochim. Biophys. Acta* 1396 (1), 105–113. doi:10.1016/s0167-4781(97)00181-4

Yamagoe, S., Yamakawa, Y., Matsuo, Y., Minowada, J., Mizuno, S., and Suzuki, K. (1996). Purification and primary amino acid sequence of a novel neutrophil chemotactic factor LECT2. *Immunol. Lett.* 52 (1), 9–13. doi:10.1016/0165-2478(96)02572-2

Yin, Z., Yan, X., Wang, Q., Deng, Z., Tang, K., Cao, Z., et al. (2020). Detecting prognosis risk biomarkers for colon cancer through multi-omics-based prognostic analysis and target regulation simulation modeling. *Front. Genet.* 11, 524. doi:10.3389/fgene.2020.00524

Yoo, H. J., Hwang, S. Y., Choi, J. H., Lee, H. J., Chung, H. S., Seo, J. A., et al. (2017). Association of leukocyte cell-derived chemotaxin 2 (LECT2) with NAFLD, metabolic syndrome, and atherosclerosis. *PLoS One* 12 (4), e0174717. doi:10.1371/journal.pone.0174717

Zhang, Y., and Zhang, Z. (2020). The history and advances in cancer immunotherapy: Understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell. Mol. Immunol.* 17 (8), 807–821. doi:10.1038/s41423-020-0488-6

Zhu, S., Bennett, S., Li, Y., Liu, M., and Xu, J. (2022). The molecular structure and role of LECT2 or CHM-II in arthritis, cancer, and other diseases. *J. Cell. Physiol.* 237 (1), 480–488. doi:10.1002/jcp.30593

frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 16 September 2022
DOI 10.3389/fphys.2022.983177

Check for updates

OPEN ACCESS

EDITED BY
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

REVIEWED BY
Malgorzata Domino,
Warsaw University of Life Sciences, Poland
Shier Nee Saw,
University of Malaya, Malaysia

*CORRESPONDENCE
Huifang Wang,
kuangwhf@163.com

SPECIALTY SECTION
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

RECEIVED 30 June 2022
ACCEPTED 02 September 2022
PUBLISHED 16 September 2022

CITATION
Wang Y, Li X, Wei N, Liu Y, Liu X, Sun R, Huang C, Yao B and Wang H (2022), Automated measurement of endometrial peristalsis in cine transvaginal ultrasound images. *Front. Physiol.* 13:983177. doi: 10.3389/fphys.2022.983177

COPYRIGHT
© 2022 Wang, Li, Wei, Liu, Liu, Sun, Huang, Yao 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.

Automated measurement of endometrial peristalsis in cine transvaginal ultrasound images

Yue Wang[1], Xiaokun Li[1], Niya Wei[1], Yuanxi Liu[2], Xinting Liu[1], Ruijie Sun[1], Chan Huang[2], Bin Yao[2] and Huifang Wang[1]*

[1]Department of Ultrasound, Peking University Shenzhen Hospital, Shenzhen, China, [2]Shenzhen Wisonic Medical Technology Co., Ltd., Shenzhen, China

Objectives: Endometrial peristalsis (EP) in non-pregnant uterine can be assessed by visual assessment of transvaginal ultrasound (TVUS). However, visual assessment is subjective, and the outcome depends on the sonographers and video analysts. This study aimed to create a newly developed automatic analysis algorithm for measuring the EP compared to visual assessment.

Methods: A retrospective analysis was performed using the datasets from *in vitro* fertilization and embryo transfer (IVF-ET), who underwent the evaluation of EP by TVUS within 5 days prior to transplantation. 158 cine TVUS images were used to develop the automated analysis algorithm, and 37 cine TVUS images were evaluated by both visual and automated analysis algorithms. The algorithm was developed by applying the optical flow technology and enabled objective analysis of the number, direction, and intensity of EP.

Results: The number of peristaltic waves counted by visual assessment was 4.2 ± 2.3 (mean ± standard deviation) and 4.1 ± 2.1 for doctors one and two, respectively. The number of waves counted with the algorithm was 3.6 ± 2.1 at first evaluation and 3.7 ± 2.0 at repeated evaluation. A significant difference was found between the algorithm count and visual assessment (p = 0.001, 0.002, 0.003, 0.008). The ICC values for algorithm versus manuals ranged from 0.84 to 0.96 and 0.87 to 0.96. The numbers of the cervix-to-fundus (CF), fundus-to-cervix (FC), and both cervix-to-fundal and fundus-to-cervix (CF + FC) directions of EP counted by the algorithm were 50, 52, and 32, respectively. The numbers counted by visual assessment were 43, 45, and 46, respectively. The number of EP was the same in 87% of the two algorithm counts. The number was lower between the algorithm and visual analysis (79% with complete agreement). The EP intensity assessed by the algorithm was 2.6 ± 1.1, and the peristalsis velocity was 0.147 (0.07) mm/s.

Conclusion: The fully automated analysis algorithm can be used to quantify uterine peristalsis comparable to visual assessment.

KEYWORDS

endometrial peristalsis, automated analysis algorithm, *in vitro* fertilization and embryo transfer, optical flow technology, transvaginal ultrasonography

1 Introduction

Endometrial peristalsis (EP) is a stripping movement of the endometrium caused by subtle, wave-like contractions of the sub-endometrial myometrium (Pinto et al., 2015). Peristaltic frequency, direction, and intensity vary according to the menstrual cycle phases under hormonal variations (Wray et al., 2014; Young, 2016). EP plays essential roles in sperm transportation, menstrual discharge, and embryo implantation, which favor pregnancy and the early development of embryos. It has been reported that any change in the velocity and direction of EP compared to its typical characteristics may lead to infertility or pregnancy failure (Kuijsters et al., 2017; Soares et al., 2019). Various diagnostic technologies, such as intrauterine pressure measurement (IUPs), magnetic resonance imaging (MRI), and transvaginal ultrasonography (TVUS), have been introduced to investigate the EP. MRI is costly and not readily available. In the case of IUPs, a significant drawback is that the irritation induced by an intrauterine device may interfere with physiological contraction characteristics, which causes discomfort for patients and makes routine use impractical (Kuijsters et al., 2017; Liu et al., 2018). TVUS, considered a non-invasive, cost-effective, and safe approach for measuring EP, is currently the most appealing method for evaluating EP (Huang et al., 2018; Kuijsters et al., 2019). However, visual assessment is subjective, and the outcome depends on the sonographers and video analysts. The necessary knowledge and skills of a doctor, and thus the need for training and appropriate qualifications (not routinely held at the basic stage of education), are necessary for assessing EP. In addition, the observation and interpretation of EP were too time-consuming to be used in daily practice, even for experienced sonographers (Kuijsters et al., 2019). To overcome these disadvantages, automated analysis of EP in TVUS videos could be a solution. However, the EP is slow and sporadic, different from those shown by cardiac contractility, is regular and distinct, and is not easily assessed automatically. This study aimed to evaluate EP by an automated technique that enables objective analysis using a newly developed automatic analysis algorithm based on optical flow technology and then to compare these results with those from traditional visual assessment by TVUS findings.

2 Materials and methods

2.1 Study population

A retrospective analysis was performed of 267 patients who underwent *in vitro* fertilization and embryo transfer (IVF-ET) in the reproductive medical center of Peking University Shenzhen hospital between October 2020 and December 2021. Within 5 days prior to transplantation, all patients underwent the evaluation of EP by TVUS. None of the patients had received anticholinergic medications and anti-spasticity agents. The exclusion criteria were the women with uterine pathologies such as adenomyosis, uterine anomaly, uterine fibroids, and polys. Women with intrauterine devices were also excluded. Finally, a total of 195 patients were included in the study. Recorded cine ultrasound images were extracted from a picture archiving and communication system (PACS). Of the 195 cine TVUS images extracted, 158 were used to develop the automated analysis algorithm, and 37 were evaluated by both visual and automated analysis algorithms. Ethical approval was given by the Ethics Committee of Peking University Shenzhen Hospital (No. 2022002). A waiver of informed consent has been obtained for this retrospective study.

2.2 Cine transvaginal ultrasound images acquisition

Two ultrasound machines available at our outpatient clinic were used to acquire the cine TVUS images: Resona7 (Mindray Medical Systems, Shenzhen, China), with a 2–9 MHz endovaginal volume transducer (DE10-3WU); and Voluson E8 (GE Healthcare, United States), with a 5–9 MHz endovaginal volume transducer (RIC 5-9-D). These systems had a built-in video record, and the recorded file was later digitized into an AVI/MP4 file.

A standardized scanning protocol was set up, and all the scans were performed according to the following protocol: 1) Let the patient lie in a supine position, keep the body still, and breathe normally. Any artifacts due to respiratory or intestinal movement were excluded; 2) Find the section of the uterus with the largest longitudinal section and the operator holding the probe steady; 3) Collect video data for 2 min; 4) Visual inspection of these ultrasound recordings, replayed at two times the regular speed, was independently performed by two doctors with more than 5 years of TVUS experience and 1 year of experience evaluating EP; 5) When peristalsis occurred, the algorithm and visual assessment evaluated the number and direction of peristalsis. The algorithm only evaluated the intensity and velocity of EP. The number is EP's number in a time, and the direction of EP was defined by the line connecting the cervix to the fundus. The direction of peristaltic waves was classified as cervix-to-fundus (CF), fundus-to-cervix (FC), and both cervix-to-fundal and fundus-to-cervix (CF + FC). The EP's velocity is defined as the time it takes one peristalsis wave from the beginning to the end. The velocity is the length of a path (mm) divided by the time (s) it takes for the peristalsis to complete the path. The EP's intensity is to calculate each peak on the peristaltic wave curve and generate a point with a

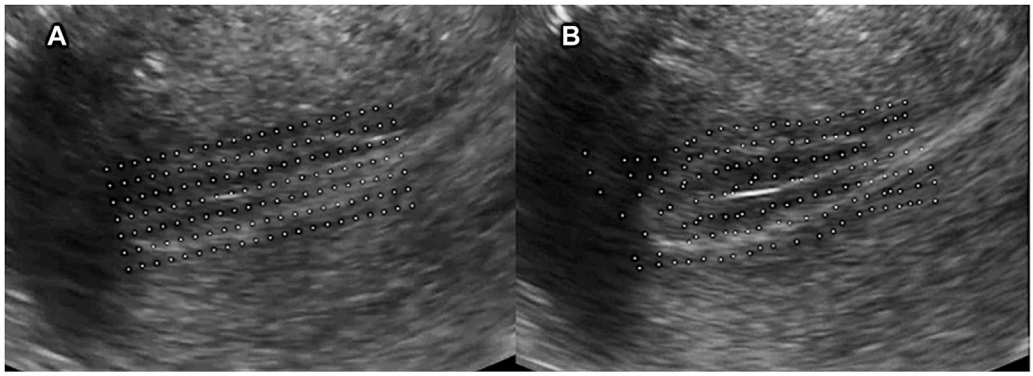


FIGURE 1
(A) Feature points generation. First, determined the rectangle that encircles the endometrium area and filled this rectangle area with aligned feature points. The feature points are equally spaced, and the interval is usually 15-pixel-length; **(B)** Displacement of feature points.

peristaltic range on the x-axis and peristaltic amplitude on the y-axis in the coordinate system.

2.3 Development of automated analysis algorithm

The algorithm had three main processing stages: inputting cine images for feature extraction, model establishment, and evaluation.

2.3.1 Algorithm establishment

2.3.1.1 Motion capturing

1) Feature points generation: when the video was imported, the rectangle was determined that encircle the endometrium area and fill this rectangle area with aligned feature points. The feature points are equally spaced, and the interval is usually 15-pixel-length. The initial coordinates of each feature point are recorded then (Figure 1A).
2) Displacement of feature points: the unit time "t" was assumed as the time duration of two adjacent video frames. Then the velocity of each feature point is simplified to its displacement between two adjacent frames, Δx. Furthermore, the new coordinates can be expressed as $x_t = x_0 + \Delta x$ (Figure 1B).
3) Temporal and spatial filtering: Due to the background noise of ultrasound, the gray value of the picture is constantly changing, which can cause an error during calculation. After coordinate data were collected from each frame, the motion information could not be obtained from this data directly. Instead, temporal and spatial smoothing processing was needed first.
4) Computing feature value: Since the peristalsis of the endometrium is a continuous motion, which has spatial continuity. It means the displace of each single feature point will not describe the whole motion and should consider this question from a macro viewpoint. If we regard points in each row as a line, we fit points whose initial position is in the same row with a straight line. And then, the peristalsis will deform the straight line into a curve. If we compare the left and right sides of Figure 2, one can observe that the line curves in the same direction with peristalsis, and the curvature is in proportion to the magnitude of peristalsis. In this condition, we can say that the curvature of those fitting curves can present the magnitude of endometrium peristalsis. We use the variation of curvature of two frames instead of curvature from a single frame since it is static. So we calculate the difference by subtracting the curvature in the frame that is five frames ahead of the curvature in the current frame, and the variation is $\Delta\tau$. (If there are less than five frames ahead of the current frame, then subtract with the first frame). In addition, we prefer to record this variation data in two parts: sign and absolute value. The absolute value presents the magnitude, and the sign stands for direction, which we will discuss later.

2.3.1.2 Motion amplifying

1) Salient motion determinant: we define the absolute value of curvature variation as the peristalsis magnitude parameter. Since the numerical value of magnitude from samples is different, we need to normalize the curvature variation data obtained from the last step to make it easier to determine the magnitude level. We use a linear normalization function that can map all the magnitude data into sections [0, 1] uniformly:

$$y_0 = \phi(|\Delta\tau|) \in [0, 1]. \quad (1)$$

And then, we can define a threshold value α, for a value y_0 larger than α can be regarded as salient peristalsis. The one less than α will be regarded as non-peristalsis. (The threshold value α

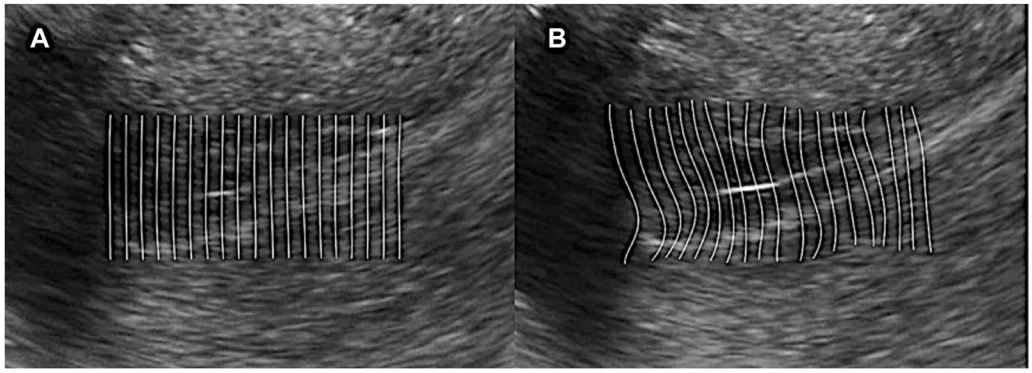


FIGURE 2
Computing feature value. **(A)** Regard points in each row as a line and fit points whose initial position is in the same row with a straight line; **(B)** Peristalsis deforms the straight line into a curve. The line curves in the same direction as peristalsis, and the curvature is in proportion to the magnitude of peristalsis.

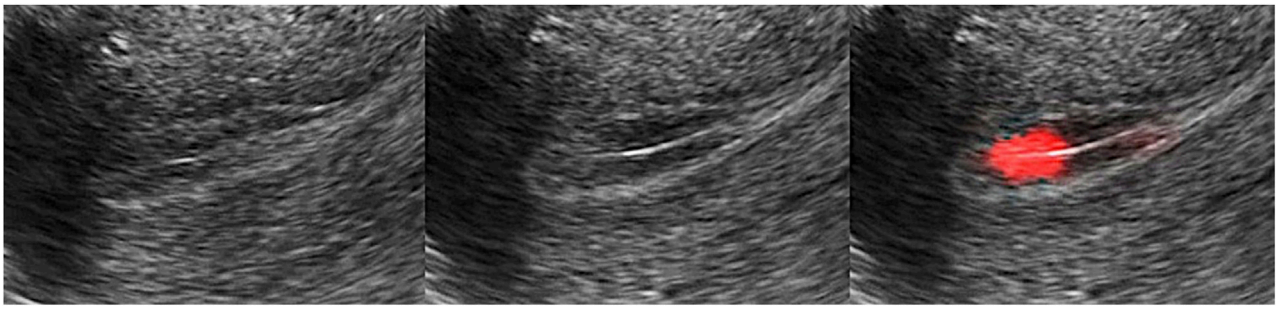


FIGURE 3
Rendering display. Multiply the color value with the rendering weight and add the value into the RGB channel of the original frame. The area with salient peristalsis was bright red, and no peristalsis kept the same grey value.

can be modified from 0 to 1, in this experiment, we choose $\alpha = 0.6$):

$$y_1 = y_0 - \alpha. \tag{2}$$

2) Rendering weight parameter: We wish to visualize the peristalsis in a color rendering way: the more salient the peristalsis is, the brighter the peristalsis area will be (High rendering weight); and vice-versa, the area without peristalsis will not be rendered (Low rendering weight). We modify the weight parameter by the Sigmoid function:

$$y_2 = sigmoid(y_1, \beta) = \frac{1}{1 + e^{-\beta \cdot y_1}}. \tag{3}$$

β controls the slope of the Sigmoid function, a larger slope means the peristalsis area will have higher weight, and the non-peristalsis area will get lower weight.

3) Rendering display: To display the final rendering results, we multiply the color value with the rendering weight and add them into the RGB channel of the original frame. We also need interpolation to the whole endometrium area since we only have the value in feature points (Figure 3). As can be seen, the area with salient peristalsis is bright red. On the contrary, the area with no peristalsis keeps the same grey value.
4) Motion graph generation: Since the diversity of the endometrium orientation in different samples, we define the left orientation (for horizontal position) and downward orientation (for vertical position) as "forward direction"; The right orientation and upward orientation correspond for "backward direction." Remember, we have recorded the sign of $\Delta\tau$, which presents the peristalsis direction: "+1" stands for "forward direction" and "−1" stands for "backward direction."

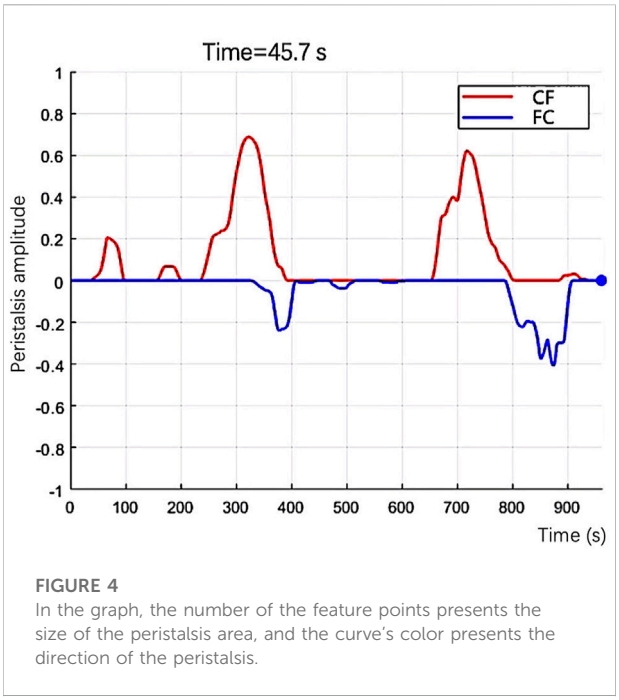


FIGURE 4
In the graph, the number of the feature points presents the size of the peristalsis area, and the curve's color presents the direction of the peristalsis.

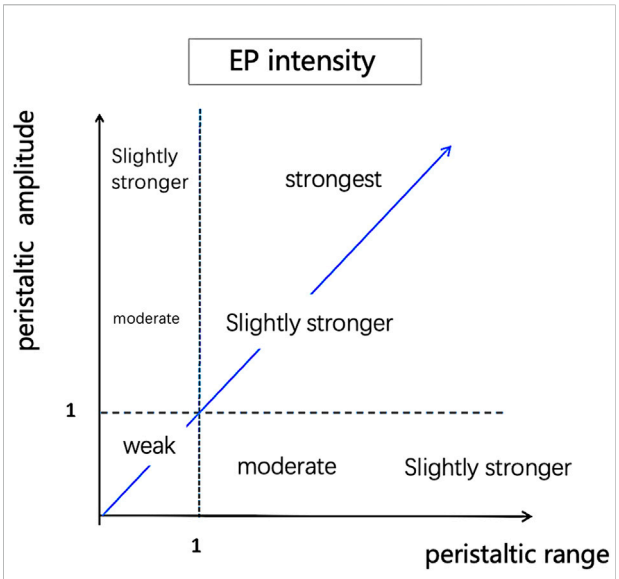


FIGURE 5
Quantitative assessment of EP intensity. Calculate each peak on the peristaltic wave curve and generate a point with peristaltic range on the x-axis and peristaltic amplitude on the y-axis in the coordinate system.

We only concern the salient peristalsis area, the feature points with amplitude parameters larger than α. We count the number of those points of 'forward direction' and 'backward direction' separately and then label them as "n_+" and "n_−." We plot "n_+" and "n_−" on the Cartesian coordinate, then obtain the motion graph. In the graph, the number of the feature points presents the size of the peristalsis area, and the curve's color presents the direction of the peristalsis (Figure 4). It is easy to observe some features of peristalsis with different modes:

A) In a one-way peristalsis motion graph, the curve presenting the "forward direction" (or "backward direction") is always above the other one, and the curve of the opposite direction will always keep zero value.
B) For the peristalsis that first move "forward" and then move "backward" (or first "backward" then "forward"), its graph has the feature that the "peak" of the curve will appear alternately.
C) If the "forward" and "backward" peristalsis happen simultaneously, then their "peak" of curves will also appear simultaneously in the graph.

2.3.2 Algorithm evaluation

(1) The established algorithm evaluated EP's number in 37 cine ultrasound datasets twice, and two physicians also evaluated the number. The repeatability of the algorithm evaluation and the agreement between the algorithm and the visual assessment was calculated.
(2) Researchers extracted 134 cine ultrasound images containing only one EP from 37 datasets. The algorithm and two sonographers evaluated EP's direction in 134 cine ultrasound images simultaneously. The consistency of EP's direction evaluation between the algorithm and the visual assessment was calculated.
(3) Quantitative assessment of EP intensity and velocity
    1) The intensity of EP was calculated at each peak on the peristaltic wave curve and generated a point with a peristaltic range on the x-axis and peristaltic amplitude on the y-axis in the coordinate system (Figure 5). The classification of EP's intensity is defined according to the following criteria:
        A) weak: the peristalsis wave range and peristalsis amplitude are both less than 1;
        B) moderate: either peristalsis wave range or peristalsis amplitude is greater than 1;
        C) strong: both the peristalsis wave range and amplitude are greater than 1;
    2) Even if the peristaltic range is the same, some waves are fast while others are slow. We use the ratio of wave peak to wavelength (i.e., the ratio of ordinate to abscissa) as the indicator of peristalsis velocity (Figure 6).

2.4 Statistical analysis

Continuous data were expressed as mean plus/minus standard deviation, and normal distribution was tested using the Shapiro-Wilk test. The EP's numbers counted by the

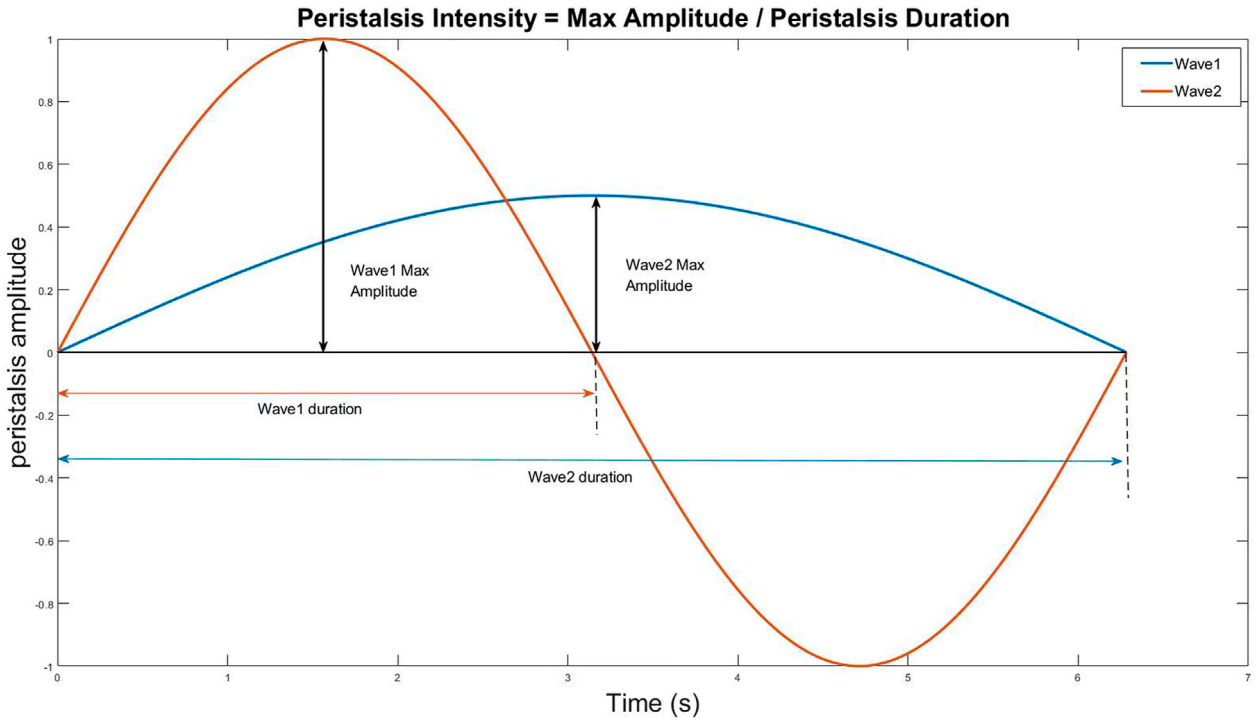


FIGURE 6
The ratio of wave peak to wavelength (i.e., the ratio of ordinate to abscissa) as the indicator of peristalsis velocity.

algorithm and visual assessment were compared with the Wilcoxon signed-rank test. The concordance was computed using the intraclass coefficient correlation (ICC). The repeatability of two algorithm counts for one video was evaluated using ICC; ICC evaluated the agreement between algorithm and visual analysis; ICC evaluated the inter-reader agreement between two visual assessments. ICC of less than 0.20 denotes poor repeatability, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, and 0.81–1.00 excellent repeatability (Youssef et al., 2016). All statistical analyses were performed using SPSS Statistics 26 (IBM SPSS Statistics for Mac, Version 26.0).

3 Results

1. In Supplementary Video S1 was an example of the automated analysis algorithm that evaluated EP. The red lines were the peristalsis from the cervix to the fundus, and the blue lines were the fundus to the cervix. The video shows three peristalsis waves from the cervix to the fundus and two from the fundus to the cervix.

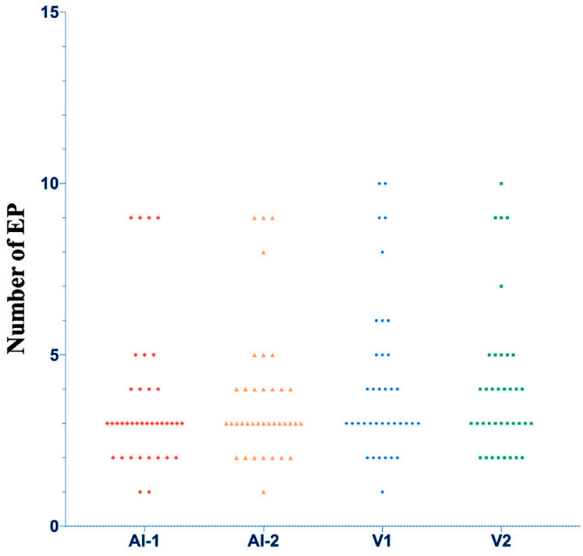


FIGURE 7
The distributions of the number of EP. AI, Artificial intelligence algorithm; V, visual analysis.

TABLE 1 Agreement in number of endometrial peristaltic waves counted by different methods.

Methods	Mean (and standard deviation)	ICC
Algorithm evaluation 1st	3.6 (2.1)	
Algorithm evaluation 2nd	3.7 (2.0)	
Visual reader 1	4.2 (2.3)	
Visual reader 2	4.1 (2.2)	
Algorithm evaluation 1st vs.2nd		0.97
Visual reader 1 vs. 2		0.96
Algorithm evaluation 1st vs. visual reader 1		0.91
Algorithm evaluation 1st vs. visual reader 2		0.93
Algorithm evaluation 2nd vs. visual reader 1		0.91
Algorithm evaluation 2nd vs. visual reader 2		0.94

TABLE 2 Direction of endometrial peristaltic waves evaluated by different methods.

	The same number	Percentage in total (%)
Algorithm evaluation 1st vs. 2nd	116	0.87
Algorithm evaluation 1st vs. visual reader 1	106	0.79
Algorithm evaluation 1st vs. visual reader 2	97	0.72
Algorithm evaluation 2nd vs. visual reader 1	97	0.72
Algorithm evaluation 2nd vs. visual reader 2	88	0.65
Visual reader 1st vs. 2nd	89	0.66

2. Repeatability and consistency of EP's number counts

The algorithm and visual assessment analyzed a total of 37 cine ultrasound images. The datasets of EP's number did not conform to a normal distribution according to a Shapiro-Wilk test (p = 0.000). The mean (and standard deviation) of the EP's number counted by visual assessment was 4.2 ± 2.3 and 4.1 ± 2.2 for doctors one and two, respectively. The EP's number counted with the algorithm was 3.6 ± 2.1 at first evaluation and 3.7 ± 2.0 at repeated evaluation. A significant difference was found between the algorithm counts and visual assessments (p = 0.001, 0.002, 0.003, 0.008). The distributions of the EP's number are shown in Figure 7.

The repeatability of the two algorithm counts was excellent, with an ICC value of 0.97 (p = 0.000). The number of five datasets was different; among them, three datasets had a difference of 1 peristalsis wave, and two datasets had a difference of 2 peristalsis waves. The algorithm counts also showed excellent agreement with the visual assessment of docotor1 (ICC values ranging from 0.84 to 0.96, p = 0.000) and doctor 2 (ICC values ranging from 0.87 to 0.96, p = 0.000). The inter-reader agreement between doctors 1 and 2 was excellent (ICC values ranging from 0.92 to 0.98, p = 0.000). ICCs value is summarized in Table 1.

3. EP direction assessment

A total of 134 cine ultrasound images containing only one EP were analyzed by algorithm evaluation and visual assessment. The number of cine ultrasound images with the direction of EP classified into CF, FC, and CF + FC per method is shown in Table 2. Mixed CF + FC direction was observed in 24% of cine ultrasound images by algorithm, while this pattern was slightly less frequent in visual assessment (22%). In the algorithm, the number of same-direction EP was the same in 87% of the two algorithm counts. The ratio was lower between the algorithm evaluation and visual assessment (79%) and between the two sonographers (66%) (Table 2).

4. The mean EP intensity assessed by the algorithm was 2.6 ± 1.1, and the mean peristalsis velocity was 0.147 (0.07) mm/s.

4 Discussion

The developed algorithm could automatically measure the number, direction, velocity, and intensity of EP in cine ultrasound images. The results indicated that the algorithm is reliable, objective, and reproducible for measuring EP. The study demonstrated that the algorithm had good repeatability in evaluating EP's number. The number was precisely the same between the two evaluations in 32 out of 37 cine ultrasound images, and the remaining five showed only one or two differences in the repeated evaluation. The results also showed that the algorithm evaluation was in close agreement with the visual assessment. The number recognized by the algorithm was less than that recognized by visual assessment (3.6 vs. 4.2), and the difference is statistically significant. By analyzing the cine ultrasound images, it was found that most differences occurred in the video with the CF + FC peristalsis wave. The possible reason might be that the sensitivity of vision to time resolution is inferior to the algorithm. The visual evaluated CF + FC as two EPs while the algorithm as one. The algorithm evaluation of EP direction was consistent with the visual assessment. In addition to the common CF, FC, and CF + FC, the EP direction also has the following conditions: 1) Peristalsis starts in the middle of the uterine corpus and then peristalsis to the uterine fundus and cervix at the same time; 2) Peristalsis co-occurs in different directions at multiple starting points; 3) The direction of peristalsis is inconsistent with the longitudinal axis of the uterus, showing irregular peristalsis. The above conditions lead to difficulty in judgment by visual assessment and algorithm evaluation. In this study, there were no datasets with 0 number of EP by visual assessment, so it was impossible to judge the advantage of the algorithm over visual assessment in spatial resolution.

This study has two noteworthy strengths. First, an assessment based on a multi-indicator approach could provide more comprehensive information for the clinical practice. Not only the number and direction but also the velocity and intensity of EP could be evaluated by the algorithm. Peristaltic waves of the same number and direction must have different physiological and physical effects on the endometrium if the peristaltic range is too extensive or the velocity is too fast. No published studies have assessed the intensity and velocity of peristalsis by ultrasound. Second, EP's four indicators are presented in coordinates, the number of peristalses was the number of waves, the direction was different colors up and down the X-axis, and the amplitude of the wave displays the intensity of the peristalsis. The velocity of peristalsis is the wave's speed in the video's coordinate system.

Although EP has been extensively studied as a factor affecting fertility since the 1990s, the assessment is currently not used as a routine examination in clinical practice, mainly due to the lack of an efficient, objective, accurate assessment method. IUPs are theoretically the most accurate and objective for determining all effects and dimensions of EP. A significant drawback of IUPs is that the device causes the patient discomfort. In addition, irritation induced by an intrauterine device may interfere with physiological contraction characteristics (Wray et al., 2014). MRI can measure the frequency of EP but not amplitude. MRI has a higher detection rate because it is more advantageous in displaying sub-endometrial wave conduction. However, MRI is expensive and time-consuming (Kido et al., 2011; Nakai et al., 2012; Watanabe et al., 2014). Tasnim's team (Tasnim et al., 2019) and Watanabe's team (Watanabe et al., 2014) investigated the number of EPs automatically assessed by MRI imaging.

Van Gestel et al. (2007) used ultrasound to evaluate EP. The study showed that the interobserver agreement among the three investigators resulted in a kappa value of 0.83, reflecting strong agreement. The study did not explore the consistency of contraction amplitude. Mori's team established a model for predicting pregnancy outcomes by ultrasound assessment of uterine motion velocity (Morizaki et al., 1989). Huang's team applied speckle tracking technology to automatically assess the velocity and direction of contraction waves (Huang et al., 2018).

Limitations of our study include its retrospective nature and the small sample size. EP was only analyzed from a methodological point of view and was not evaluated in conjunction with clinical pregnancy outcomes. Future prospective studies of EP combined with clinical pregnancy outcomes and different menstrual cycles are needed. As the algorithm is in the experimental stage, technical problems such as complex programs and motion artifacts will be solved in the future.

In conclusion, we developed the automated analysis algorithm based on optical flow technology, which can comprehensively evaluate EP's number, direction, intensity, and velocity in cine ultrasound images. The algorithm can improve the efficiency of clinical evaluation of EP and has potential application prospects.

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 Ethics Committee of Peking University Shenzhen Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YW and HW contributed to the conception and design of the study. XkL, NW, RS, and XtL organized the database. YW performed the statistical analysis. YW wrote the first draft of the manuscript. YL, CH, and BY wrote sections of the materials and methods. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This study was supported by Shenzhen Key Medical Discipline Construction Fund (No. SZXK051), Guangdong High-level hospital construction fund (No. GD2019260), and Sanming Project of Medicine in Shenzhen (No. SZSM202111011).

Conflict of interest

Authors YL, CH, and BY were employed by Shenzhen Wisonic Medical Technology Co., Ltd.

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/fphys.2022.983177/full#supplementary-material

SUPPLEMENTARY VIDEO S1
An example of the automated analysis algorithm evaluated EP. The red lines were the peristalsis from the cervix to the fundus, and the blue lines were the fundus to the cervix. The video shows three peristalsis waves from the cervix to the fundus and two from the fundus to the cervix.

References

Huang, Y., Sammali, F., Blank, C., et al. (2018). Quantitative ultrasound imaging and characterization of uterine peristaltic waves, Proceedings of the IEEE international ultrasonics symposium (IUS). Kobe, Japan, IEEE. doi:10.1109/ultsym.2018.8580218

Kido, A., Ascher, S. M., Kishimoto, K., Hahn, W., Jha, R. C., Togashi, K., et al. (2011). Comparison of uterine peristalsis before and after uterine artery embolization at 3-T MRI. *Am. J. Roentgenol.* 196 (6), 1431–1435. doi:10.2214/ajr.10.5349

Kuijsters, N. P. M., Methorst, W. G., Kortenhorst, M. S. Q., Rabotti, C., Mischi, M., and Schoot, B. C. (2017). Uterine peristalsis and fertility: Current knowledge and future perspectives: A review and meta-analysis. *Reprod. Biomed. Online* 35 (1), 50–71. doi:10.1016/j.rbmo.2017.03.019

Kuijsters, N. P. M., Sammali, F., Rabotti, C., Huang, Y., Mischi, M., and Schoot, B. C. (2019). Visual inspection of transvaginal ultrasound videos to characterize uterine peristalsis: An inter-observer agreement study. *J. Ultrasound* 23 (1), 37–44. doi:10.1007/s40477-018-00356-z

Liu, S., Zhang, Q., Yin, C., Chen, W., Chan, Q., He, J., et al. (2018). An optimised repetition time (TR) for cine imaging of uterine peristalsis on 3 T MRI. *Clin. Radiol.* 73 (7), 678. e7–678.e12. doi:10.1016/j.crad.2018.03.012

Morizaki, N., Morizaki, J., Hayashi, R. H., and Garfield, R. E. (1989). A functional and structural study of the innervation of the human uterus. *Am. J. Obstetrics Gynecol.* 160 (1), 218–228. doi:10.1016/00029378(89)90126-9

Nakai, A., Reinhold, C., Noel, P., Kido, A., Rafatzand, K., Ito, I., et al. (2012). Optimizing cine MRI for uterine peristalsis: A comparison of three different single shot fast spin echo techniques. *J. Magnetic Reson. Imaging* 38 (1), 161–167. doi:10.1002/jmri.23946

Pinto, V., Matteo, M., Tinelli, R., Mitola, P. C., De Ziegler, D., and Cicinelli, E. (2015). Altered uterine contractility in women with chronic endometritis. *Fertil. Steril.* 103 (4), 1049–1052. doi:10.1016/j.fertnstert.2015.01.007

Soares, D. M., Junior, H. W., Bittencourt, L. K., Lopes, F. P. P. L., and de Oliveira, M. A. P. (2019). The role of cine MR imaging in the assessment of uterine function. *Archives Gynecol. Obstetrics* 300 (3), 545–553. doi:10.1007/s00404-019-05222-9

Tasnim, N., Mahdi, F. P., Alam, S. B., et al. (2019). A quantification approach of uterine peristalsis propagated from the cervix and the fundus, Proceedings of the International conference on machine learning and cybernetics (ICMLC). IEEE. Shaanxi, China, doi:10.1109/icmlc48188.2019.8949268

van Gestel, I., IJland, M. M., Evers, J. L. H., and Hoogland, H. J. (2007). Complex endometrial wave-patterns in IVF. *Fertil. Steril.* 88 (3), 612–615. doi:10.1016/j.fertnstert.2006.12.050

Watanabe, K., Kataoka, M., Yano, K., Nishio, S., Umehana, M., Kido, A., et al. (2014). Automated detection and measurement of uterine peristalsis in cine MR images. *J. Magnetic Reson. Imaging* 42 (3), 644–650. doi:10.1002/jmri.24817

Wray, S., Burdyga, T., Noble, D., Noble, K., Borysova, L., and Arrowsmith, S. (2014). Progress in understanding electro-mechanical signalling in the myometrium. *Acta Physiol.* 213 (2), 417–431. doi:10.1111/apha.12431

Young, R. C. (2016). Mechanotransduction mechanisms for coordinating uterine contractions in human labor. *Reproduction* 152 (2), R51–R61. doi:10.1530/rep-16-0156

Youssef, A., Salsi, G., Cataneo, I., Martelli, F., Azzarone, C., Bellussi, F., et al. (2016). Agreement between two 3D ultrasound techniques for the assessment of the subpubic arch angle. *J. Maternal-Fetal Neonatal Med.* 30 (4), 466–470. doi:10.1080/14767058.2016.1175000

frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 10 October 2022
DOI 10.3389/fphys.2022.956254

Check for updates

OPEN ACCESS

EDITED BY
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

REVIEWED BY
Serap Aydin,
Hacettepe University, Turkey
Mengfan Li,
Hebei University of Technology, China

*CORRESPONDENCE
Libin Deng,
denglb1937@163.com
Xing Wang,
xing.wang@ncu.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

RECEIVED 30 May 2022
ACCEPTED 20 September 2022
PUBLISHED 10 October 2022

CITATION
Li M, Liu Y, Liu Y, Pu C, Yin R, Zeng Z, Deng L and Wang X (2022), Resting-state EEG-based convolutional neural network for the diagnosis of depression and its severity.
Front. Physiol. 13:956254.
doi: 10.3389/fphys.2022.956254

COPYRIGHT
© 2022 Li, Liu, Liu, Pu, Yin, Zeng, Deng 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.

Resting-state EEG-based convolutional neural network for the diagnosis of depression and its severity

Mengqian Li[1], Yuan Liu[1], Yan Liu[2], Changqin Pu[3], Ruocheng Yin[3], Ziqiang Zeng[4,5], Libin Deng[4]* and Xing Wang[6,7]*

[1]Department of Psychosomatic Medicine, The First Affiliated Hospital of Nanchang University, Nanchang, China, [2]Second Clinical Medical College, Nanchang University, Nanchang, China, [3]Queen Mary College, Nanchang University, Nanchang, China, [4]Jiangxi Provincial Key Laboratory of Preventive Medicine, School of Public Health, Nanchang University, Nanchang, China, [5]School of Public Health, Nanchang University, Nanchang, China, [6]School of Life Sciences, Nanchang University, Nanchang, China, [7]Clinical Medical Experiment Center, Nanchang University, Nanchang, China

Purpose: The study aimed to assess the value of the resting-state electroencephalogram (EEG)-based convolutional neural network (CNN) method for the diagnosis of depression and its severity in order to better serve depressed patients and at-risk populations.

Methods: In this study, we used the resting state EEG-based CNN to identify depression and evaluated its severity. The EEG data were collected from depressed patients and healthy people using the Nihon Kohden EEG-1200 system. Analytical processing of resting-state EEG data was performed using Python and MATLAB software applications. The questionnaire included the Self-Rating Anxiety Scale (SAS), Self-Rating Depression Scale (SDS), Symptom Check-List-90 (SCL-90), and the Eysenck Personality Questionnaire (EPQ).

Results: A total of 82 subjects were included in this study, with 41 in the depression group and 41 in the healthy control group. The area under the curve (AUC) of the resting-state EEG-based CNN in depression diagnosis was 0.74 (95%CI: 0.70–0.77) with an accuracy of 66.40%. In the depression group, the SDS, SAS, SCL-90 subscales, and N scores were significantly higher in the major depression group than those in the non-major depression group ($p < 0.05$). The AUC of the model in depression severity was 0.70 (95%CI: 0.65–0.75) with an accuracy of 66.93%. Correlation analysis revealed that major depression AI scores were significantly correlated with SAS scores (r = 0.508, $p = 0.003$) and SDS scores (r = 0.765, $p < 0.001$).

Conclusion: Our model can accurately identify the depression-specific EEG signal in terms of depression diagnosis and severity identification. It would eventually provide new strategies for early diagnosis of depression and its severity.

KEYWORDS

EEG, convolutional neural network, depression, severity, diagnosis

1 Introduction

Depression is a common mood disorder that has negative impacts on a patient's physical and mental health (McCarron et al., 2021; Unursaikhan et al., 2021). The clinical presentation included depressed mood, slowed thinking, and decreased willpower activity. In severe cases, patients might also develop suicidal attempts (Smith, 2014). With the continuous development of human society, the number of people with depression is increasing year by year worldwide (Moreno-Agostino et al., 2020). The World Health Organization (WHO) showed that more than 300 million people worldwide suffer from depression, and about 800,000 of them commit suicide (Levey et al., 2019). Untimely identification of depression may be one of the leading causes of this result. Therefore, early diagnosis of depression is critical.

However, the objectivity and accuracy of depression diagnosis are limited by the current diagnostic criteria for depression. Some facts must be admitted: the diagnostic technique in psychiatry has historically lagged behind other domains (Murray et al., 2021). Fortunately, this challenge is being alleviated by the application of electroencephalogram (EEG) measurement. To date, EEG has been widely used in neuroscience to get insights into brain activity (Latreille et al., 2016; Schönenberg et al., 2017; Wu et al., 2020; Dimitriadis, 2021; Simonato et al., 2021). EEG recordings benefit from shorter test times and lower prices than functional magnetic resonance imaging (fMRI), making them more suitable for diagnosing several types of mental diseases (Čukić et al., 2020). In addition to using traditional EEG images for analysis, the frequency domain features of EEG images have also been shown to be one of the most useful pragmatic markers for diagnosing depression. The frequency domain analysis realizes the conversion of the EEG signal from the time domain to the frequency domain. The frequency domain analysis results in the energy value distribution at each frequency, that is, the power value. For example, Stewart et al. (2010) found that the average alpha power difference measured in the left hemisphere and the right hemisphere in depression patients is larger than that in normal people. Compared with normal people, the left hemisphere activity of depression patients is reduced (expressed as increased alpha power). At the same time, a study suggested that the energy asymmetry of the frontal lobe alpha wave in patients with depression was more obvious to the left, and the severity of symptoms was positively correlated with laterality (Grünewald et al., 2018).

With the rise of computational psychiatry (Geng et al., 2022), EEG-based machine learning (ML) to detect illness phenotypes has attracted growing interest, which provides a theoretical basis and feasibility for disease diagnosis. Since Ahmadlou et al. (2012) initially used ML approaches to detect depression early, much relevant research has been published with promising findings, especially in depression diagnosis (Puthankattil and Joseph, 2012; Hosseinifard et al., 2013; Faust et al., 2014; Bairy et al., 2015). For example, Khodayari-Rostamabad et al. (2010) proposed a diagnostic model trained using EEG data. The model was able to differentiate between subjects with major depressive disorder, chronic schizophrenia, bipolar depression, and healthy subjects by analyzing patients' EEG data. Meanwhile, Kang et al. (2020) converted the asymmetric features of EEG signals into matrix images, used them as the input of the convolutional neural network, and obtained 98.85% accuracy in depression screening. All of the aforementioned research demonstrates that combining machine learning with EEG signaling can be an effective tool for screening depression patients.

Feature extraction and selection is an essential step in ML, which could improve the model's performance. Many researchers have proposed various feature extraction and selection methods to improve the performance of resting-state EEG-based ML in differentiating depressed patients from normal controls (Wan et al., 2019; Duan et al., 2020). Despite the advantages mentioned previous to these steps, it has drawbacks, particularly the length of training time needed to obtain reliable classification results. Because of this, more and more researchers have applied deep learning (DL), especially convolutional neural networks (CNNs) (Zhang et al., 2021), to disease diagnosis (Morabito et al., 2017; Ortiz et al., 2021). CNN was a new disease detection model with adaptive learning capability. Its advantage was that, without the need to manually select features, it could shorten the experimental process. Acharya et al. (2018) initially used the CNN to identify the resting-state EEG data on normal and depressed patients with good classification performance, which has attracted significant attention. Many studies have considered that the CNN could be used as a clinically effective computer-aided diagnosis (CAD) system for depression (Kang et al., 2020; Uyulan et al., 2021).

However, the significant limitations to previous studies were two aspects. First, most studies are too strict on resting-state EEG signal preprocessing, which leads to a large number of valuable missing resting-state EEG signals and may overestimate the accuracy of the model. More realistic and high-quality data would help CNN identify the full range of depression in a more clinically meaningful and generalizable way. Second, most studies have only discriminated normal individuals from depression patients without predicting depression severity. It was reported that the severity of depression determined the symptoms, manifestations, and prognosis of the disease (Zimmerman et al., 2018). The clinical potential of deep learning has been undermined by the lack of external validation of models driven by a single dataset and by the increasing use of opaque decision-making frameworks. Therefore, overcoming these challenges is critical to harness the potential of deep learning algorithms to improve patient care and pave the way for interpretable, evidence-based machine learning in the medical imaging community.

It is worth noting that previous studies have been strict with the preprocessing of image information when using EEG modeling (raw EEGs were preprocessed and retained only the image features of depression that had been identified in previous studies). Although this strategy maximizes model accuracy, it also misses the opportunity to discover new depression-specific EEG features. Therefore, we chose to simplify the EEG processing conditions in an attempt to obtain new depression-specific EEG signatures.

Our study ensured realistic and high-quality data by reducing EEG preprocessing and adding EEG screening. In addition, this study would also predict the severity of depression to provide a reliable basis for achieving an accurate diagnosis or clinical decision-making. We chose to simplify the EEG processing conditions in an attempt to obtain new depression-specific EEG features. Using this strategy, we were able to identify new depression-specific EEG signatures in subsequent studies by using techniques such as 'deconvolutional neural networks' and to further explore the physiological impact of depression. Our study provides new strategies for the clinical diagnosis of depression.

2 Manuscript formatting

2.1 Methods

2.1.1 Subjects

A total of 41 depressed patients hospitalized in The First Affiliated Hospital of Nanchang University from September 2020 to April 2021 were selected as the depression group. Meanwhile, 41 healthy people were selected as the healthy control group. Enrollment criteria for the depression group included age 16–65 years, and the patients reached the diagnostic criteria for depression using the International Classification of Diseases, 10th edition (ICD-10). Exclusion criteria included the following: prior diagnosis of somatic disorders, bipolar disorder, schizophrenia, and other psychiatric disorders. Enrollment criteria of the healthy control group included age 16–65 years, and none met the diagnostic criteria for any psychiatric disorder using ICD-10. Exclusion criteria included the following: prior diagnosis of any somatic disorders. Informed verbal consent was obtained for all participants. Moreover, the research was approved by the Research Ethics Board at The First Affiliated Hospital of Nanchang University (approval number: 2022CDYFYYLK(06-030)). All subjects were asked to complete the Self-Rating Depression Scale (SDS) to assess the severity of depression.

2.1.2 Sample for model training/testing

Of the 62 eligible study subjects, there were 30 healthy people (without brain disease), 16 with mild to moderate depression, and 16 with severe depression (the diagnosis of mild, moderate, and moderate depression is performed using the SDS) Each patient's EEG can be cut into 60 images that meet the requirements. Thus, a total of 1800 images of healthy people, 960 images of patients with mild to moderate depression, and 960 images of patients with severe depression were taken.

Both models use a 10-fold crossover method to divide the image dataset into training and test sets in a ratio of 8:2. In the "Distinguishing Depression Model" (patients with mild to moderate depression and patients with severe depression are divided into a whole), 1440 images of healthy people and 1536 images of patients with depression are included in the training set, and 360 images of healthy people and 384 depression images were included in the test set.

In the "Model for Distinguishing Depression Severity," 768 images of patients with non-major depression (mild to moderate depression) and 768 images of patients with major depression (severe depression) were included in the training set; 192 images of patients with non-major depression and 192 images of patients with major depression were included in the test set.

2.1.3 Study design

In this study, resting-state EEG signals were collected from the depression group and the healthy control group using the Nihon Kohden EEG-1200 system. A clinical questionnaire survey and disease duration were conducted in the depression group. Gender and age were recorded for all participants. The CNN was used as a classification prediction model for depression. A flowchart of the study is listed as follows (seen in Figure 1).

Due to the characteristics of CNN and the technical limitations to the research group, we cannot use untransformed EEG signals for training (CNN technology uses image data for analysis to obtain different groups of image features and achieve classification). In this study, we selected qualified EEGs of patients who met the requirements and obtained their 600s EEG signals. According to the setting of taking one image in 10 s, we could obtain 19-channel images of a single patient.

2.1.3.1 Measurement

The measurements contained the following four parts: the Self-Rating Anxiety Scale, Self-Rating Depression Scale, Symptom Check-List-90, and the Eysenck Personality Questionnaire.

Self-Rating Anxiety Scale (SAS): The questionnaire is a validated tool for screening anxiety disorders with good reliability and validity (Ding and Yao., 2020; Dunstan et al., 2020). The SAS consisted of 20 items with a total score of 100. A higher score reflects a severer anxiety symptom.

Self-Rating Depression Scale (SDS): This is a widely used measure to screen for depression in clinical settings (Dunstan

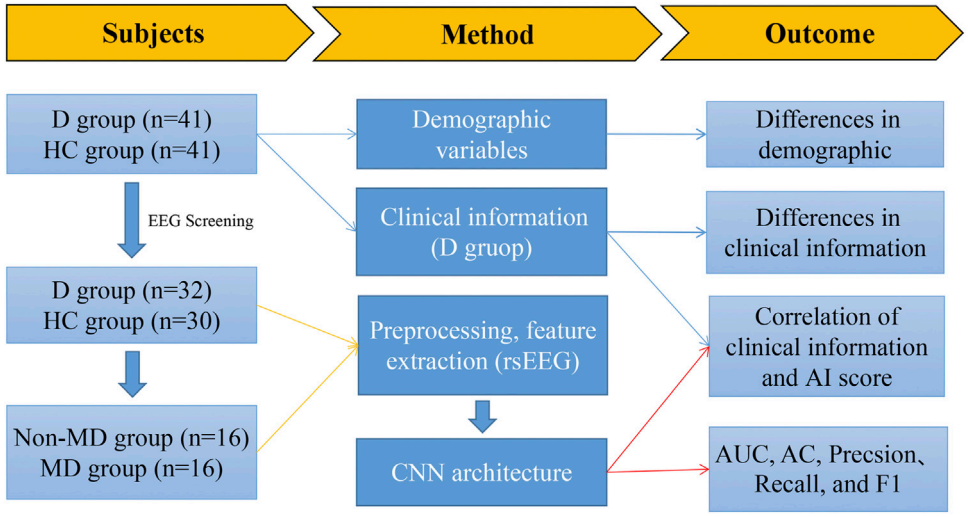


FIGURE 1
Flowchart illustrating the main content of this study. Note: D group, depression group; HC group, health control group; non-MD group, non-major depression group; MD group, major depression group.

et al., 2019; Ding and Yao., 2020). The SDS consisted of 20 items. A higher score reflects a severer depression symptom. In this study, we defined a major depression group as having a SDS score≥73. Conversely, the others were defined as non-major depression groups.

Symptom Check-List-90 (SCL-90): This scale is one of the most widely used mental health measures and has high reliability and validity (Derogatis et al., 1973). It consists of 90 items that could be divided into the symptom dimensions of somatization, obsessive-compulsive disorder, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Mental disorder is determined by total score ≥160 points or >2 points for any factor.

Eysenck Personality Questionnaire (EPQ): This is a common self-report personality questionnaire with high validity and reliability. The scale comprised 88 items summarized as extraversion (E), neuroticism (N), psychoticism (P), and lying scales (L) (EPQ, Chinese version) (Gong, 1984). The higher the score, the more likely the patient has the personality traits listed on the scale (Chen et al., 2022).

2.1.3.2 EEG recording

Here, 10 min of resting-state EEG signals were acquired in the eye-closed (EC) conditions according to the 10–20 electrode placement standard. EEG signals were recorded in the frontal (FP1, FP2, F3, F4, F7, F8, and Fz), temporal (T3, T4, T5, and T6), parietal (P3, P4, and Pz), occipital (O1 and O2), and central (C3, C4, and Cz) regions. In addition, we used the bilateral mastoids (A1 and A2) as reference electrodes. EEG signals were collected from 19 channels at a sample rate of 500 Hz. These signals were filtered with a 0.5 Hz–50 Hz bandpass filter and an additional 50-Hz notch filter. The impedance of all electrodes was within a reasonable range.

2.1.3.3 EEG preprocessing and selecting

Signal preprocessing of resting-state EEG was performed through the public MATLAB toolbox EEGLAB (Delorme and Makeig, 2004). The steps for selecting resting-state EEG signals with more than 50% effective segments are as follows: first, EEG signals were filtered offline using an FIR bandpass filter (0.5–50 Hz), and then, a notch filter was applied to remove the power frequency interference at 50 Hz. Second, EEG signals were segmented into 2 s long epochs with 300 epochs. Third, bad electrodes were removed with subsequent interpolation. Fourth, independent component analysis (ICA) was applied to identify and remove the eye blink artifacts. Fifth, epochs containing EEG amplitudes that were greater than ±70 uV were rejected automatically. Finally, the effective EEG signals were preprocessed by the aforementioned step 1 and were saved in standard EDF format for future analysis (seen in Figure 2).

2.1.3.4 EEG feature extraction

EEG features were extracted using Python and the MNE toolbox (version 0.23.4) (www.martinos.org/mne). The main steps included the following: 1) the number of electrodes and the amplitude information of each electrode were extracted. 2) The time domain features of the EEG signal were generated by aggregating the amplitude information of all electrodes in the EEG signal. 3) The time domain features

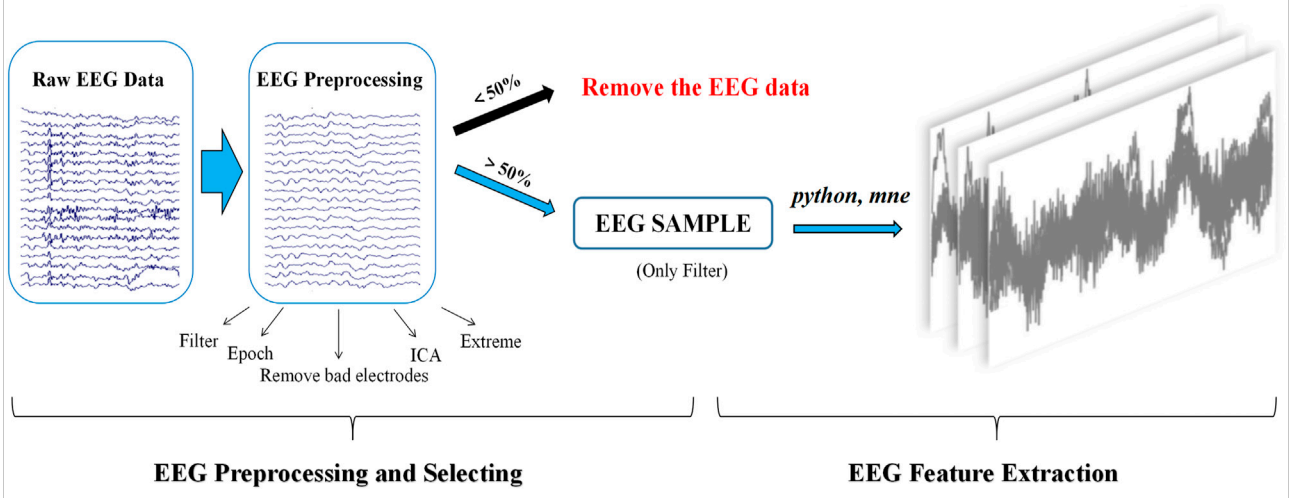


FIGURE 2
Block diagram for EEG preprocessing, selecting, and feature extraction.

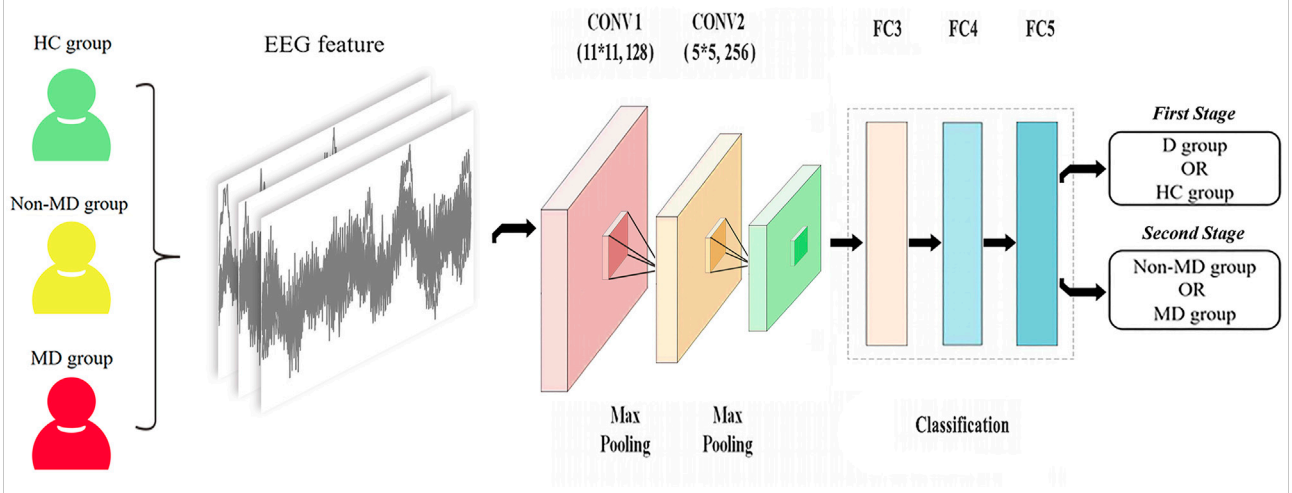


FIGURE 3
Illustration of the EEG-based CNN for diagnosis of depression and its severity. Note: 1) D group, depression group; HC group, health control group; non-MD group, non-major depression group; MD group, major depression group. 2) CONV1 (11*11, 128): 11*11 represents the size of the convolution kernel, representing height and width, respectively, and 128 represents the number of convolution kernels; CONV2 (5*5, 256): 5*5 represents the size of the convolution kernel, representing height and width, respectively, and 256 represents the number of convolution kernels.

were then segmented by a window slide (length of time duration window: 10 s) with a window step of 10 s and no overlap between time duration windows, resulting in a resting-state EEG feature map. For each case of resting-state EEG signals that met the inclusion criteria, 60 images of resting-state EEG features were generated in the PNG format (seen in Figure 2).

2.1.3.5 CNN architecture

This study adopts a deep learning model using a CNN. The model consisted of three convolutional, two maximum pooling, and three fully connected layers. The convolutional layer extracts the EEG signal features, each with 128, 256 output units. The pooling layer can reduce redundant information. Dropout and L2 regularization were added in the fully connected layers to help

prevent overfitting. The network input was a 400 × 400-pixel image. The training set was calculated using stochastic gradient descent. The validation set was used to hyper-parameterize all the networks to finalize the optimal learning rate of 0.001, and the batch size was 32. A total of 32 images were put into the training each time, with a training count of 20, to build the CNN model. The test set was put into the model for testing, and each image was calculated to obtain an AI score from 0 to 1. In fact, each image can obtain the AI score in the model (seen in Figure 3). It illustrates the EEG-based CNN for depression diagnosis and its severity. In our study, the probability of the image being diagnosed with major depression in the depression group was defined as the image's major depression AI score.

2.1.3.6 Performance evaluation of the CNN

Data were partitioned into three sets (64/16/20) to obtain training, validation, and test sets using the 10-fold cross-validation strategy. The training set was used for model training, while the validation set was used for external validation of the model. The test set was used for evaluating the final performance of the trained model. In addition, we also used the area under the curve (AUC), accuracy (AC), precision, recall, and F1-score (F1) to evaluate the performance of the proposed model. The evaluation metrics are defined as follows:

$$\mathbf{AC} = \frac{(\boldsymbol{TP} + \boldsymbol{TN})}{(\boldsymbol{TP} + \boldsymbol{FN} + \boldsymbol{TN} + \boldsymbol{FP})} \quad (1)$$

$$\mathbf{Precision} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP} + \boldsymbol{FP})} \quad (2)$$

$$\mathbf{Recall} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP} + \boldsymbol{FN})} \quad (3)$$

$$\mathbf{F1} = \frac{(\mathbf{2} \times \boldsymbol{TP})}{(\mathbf{2} \times \boldsymbol{TP} + \boldsymbol{FP} + \boldsymbol{FN})} \quad (4)$$

where TP means true positive, TN means true negative, FP means false positive, and FN means false negative.

2.1.3.7 Obtaining the optimal model parameters

In the training set, we use the 10-fold cross-validation method to divide the data into an internal training set and an internal validation set by 8:2. The internal training set participates in the training of the model, and the internal validation set is used to initially evaluate the model's performance. In this study, we set the number of iterations to 20, and the optimal learning rate is 0.0001. At the same time, during the training process, we can output the accuracy of the internal training set and internal validation set after each iteration. To avoid overfitting, the model with the highest accuracy of the two was chosen, which is regarded as the optimal model (Kiliç et al., 2022).

3.1.4 Statistical analysis

The measurement data conforming to the normal distribution were expressed as means ± standard deviation. Otherwise, the data were expressed as the median (lower and upper quartiles). The independent samples *t*-test or Mann–Whitney *U* test was conducted for the intergroup comparisons accordingly. The count data were expressed as rates and were compared using the chi-squared test. Then, Spearman's correlation analysis was used to explore the correlation between clinical characteristics and the major depression AI score. $p < 0.05$ was considered to be statistically significant. All data analyses were performed by SPSS 26.0 software.

2.2 Results

2.2.1 Demographic characteristics and clinical differences

A total of 82 subjects were included in this study, with 41 in the depression group and 41 in the healthy control group. In the depression group, 31.71% were male, and the median age was 21 years (17–37), with a median disease duration of 1 year (0.38–3). The mean scores of the SAS and SDS were 61.13 ± 2.14 and 69.97 ± 13.39, respectively. The mean scores of the SCL-90 subscales were as follows: somatization (2.47 ± 1.00), obsessive-compulsive disorder (2.98 ± 0.97), interpersonal sensitivity (2.87 ± 1.16), depression (3.21 ± 1.14), anxiety (3.04 ± 1.09), hostility (2.65 ± 1.28), phobic anxiety (2.68 ± 1.15), paranoid ideation (2.50 ± 1.17), and psychoticism (2.60 ± 1.07). The following are the mean scores of the EPQ subscales: E (40.03 ± 14.28), N (62.72 ± 11.69), P (54.37 ± 9.74), and L (41.92 ± 9.53). In the healthy control group, 36.59% were male, and the median age was 28 years (24–47.5). The median age for the healthy control group was higher than that for the depression group ($p < 0.05$). There was no significant difference in gender between the two groups ($P > 0.05$) (seen in Table 1).

The depression group consists of 22 patients with non-major depression and 19 patients with major depression. The SDS, SAS, SCL-90 subscales, and N score were significantly higher in the major depression group than those in the non-major depression group ($p < 0.05$), whereas for E, P, and L, no difference existed among groups. In addition, the median age for the non-major depression group was higher than that for the major depression group ($p < 0.05$). There was no significant difference in gender between the two groups ($P > 0.05$) (seen in Table 2).

2.2.2 Resting-state EEG screening

In the depression group, the resting-state EEG data met the inclusion criteria in 32 cases, with an inclusion rate of 78.05%, and the number of valid segments per EEG data was 231.33 ± 47.74. In the healthy control group, their resting-state EEG data met the inclusion criteria in 30 cases, with an inclusion rate of 73.17% and a valid number of segments per EEG data of 225.61 ± 41.29.

TABLE 1 Clinical and demographic characteristics in the depression group and the healthy control group.

	D group (n = 41)	HC group (n = 41)	$\chi 2/Z$	P
Gender, male (%)	13 (31.71)	15 (36.59)	0.22	0.64
Age (years)	21.00 (17.00, 37.00)	28.00 (24.00, 47.50)	2.49	0.01
Disease duration (years)	1.00 (0.38, 3.00)	—	—	—
SDS (mean, SD)	69.97 ± 13.39	—	—	—
SAS (mean, SD)	61.13 ± 2.14	—	—	—
SCL-90 (mean, SD)				
Somatization	2.47 ± 1.00	—	—	—
Obsessive-compulsive disorder	2.98 ± 0.97	—	—	—
Interpersonal sensitivity	2.87 ± 1.16	—	—	—
Depression	3.21 ± 1.14	—	—	—
Anxiety	3.04 ± 1.09	—	—	—
Hostility	2.65 ± 1.28	—	—	—
Phobic anxiety	2.68 ± 1.15	—	—	—
Paranoid ideation	2.50 ± 1.17	—	—	—
Psychoticism	2.60 ± 1.07	—	—	—
EPQ (mean, SD)				
E	40.03 ± 14.28	—	—	—
N	62.72 ± 11.69	—	—	—
P	54.37 ± 9.74	—	—	—
L	41.92 ± 9.53	—	—	—

D group, depression group; HC, group, health control group; E, extraversion; N, neuroticism; P, psychoticism; L, lying scales.

TABLE 2 Clinical and demographic characteristics in the non-major depression group and the major depression group.

	Non-MD group (n = 22)	MD group (n = 19)	$\chi 2/Z/t$	P
Gender, male (%)	7 (31.82)	6 (31.58)	0	0.99
Age (years)	25.00 (20.00, 42.00)	18.00 (16.00, 26.00)	−2.51	0.01
Disease duration (years)	1.00 (0.25, 5.25)	1.00 (0.50, 2.00)	−0.07	0.95
SDS (mean, SD)	63.75 (56.25, 69.06)	78.75 (76.25, 85.00)	−5.47	<0.001
SAS (mean, SD)	53.47 ± 10.92	70.00 ± 11.16	−4.79	<0.001
SCL-90 (mean, SD)				
Somatization	2.07 ± 0.87	2.95 ± 0.95	−3.10	0.004
Obsessive-compulsive disorder	2.53 ± 0.92	3.51 ± 0.73	−3.73	0.001
Interpersonal sensitivity	2.28 ± 1.01	3.56 ± 0.94	−4.17	<0.001
Depression	2.53 ± 0.98	4.01 ± 0.72	−5.45	<0.001
Anxiety	2.50 ± 0.98	3.67 ± 0.87	−4.04	<0.001
Hostility	2.27 ± 1.39	3.10 ± 0.98	−2.17	0.04
Phobic anxiety	2.18 ± 1.05	3.26 ± 1.00	−3.39	0.002
Paranoid ideation	1.91 ± 0.90	3.18 ± 1.09	−4.10	<0.001
Psychoticism	2.04 ± 0.82	3.26 ± 0.96	−4.41	<0.001
EPQ (mean, SD)				
E	43.30 ± 15.89	36.25 ± 11.42	1.61	0.12
N	58.76 ± 12.76	67.30 ± 8.50	−2.48	0.02
P	53.14 ± 10.76	55.78 ± 8.48	−0.86	0.39
L	41.80 ± 11.44	42.05 ± 7.00	−0.08	0.94

Note: non-MD, group, non-major depression group; MD, group, major depression group; E, extraversion; N, neuroticism; P, psychoticism; L, lying scales.

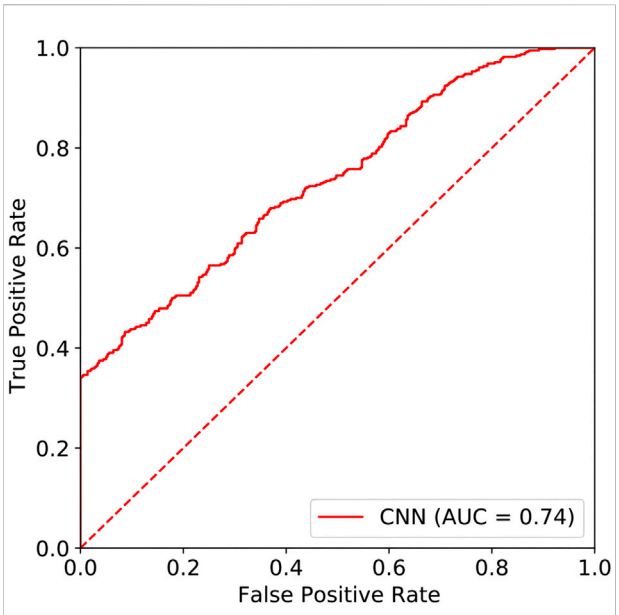


FIGURE 4
ROC of the EEG-based CNN in depression diagnosis.

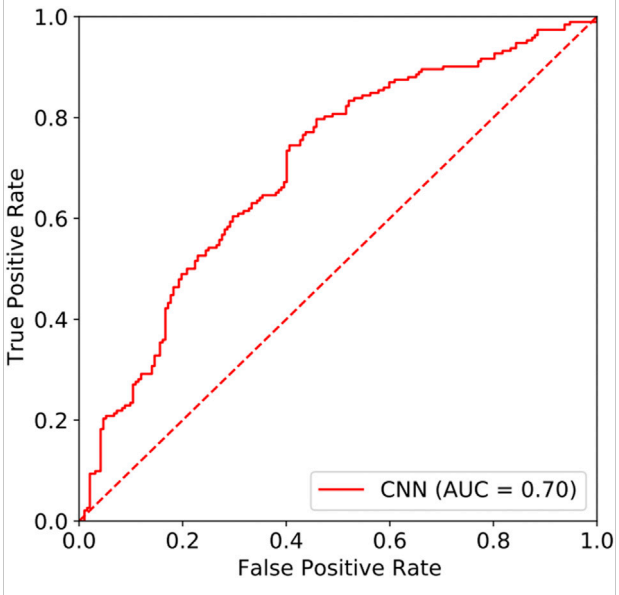


FIGURE 5
ROC of the EEG-based CNN in depression severity.

In the non-major depression group, the resting-state EEG data met the inclusion criteria in 16 cases, with an inclusion rate of 72.73% and a valid number of segments per EEG data (238.69 ± 45.26). In the major depression group, the resting-state EEG data met the inclusion criteria in 16 cases, with an inclusion rate of 84.21% and a valid number of segments per EEG data (223.06 ± 45.81).

2.2.3 CNN performance

In this study, the area under the curve (AUC) of the resting-state EEG-based CNN in depression diagnosis was 0.74 (95%CI: 0.70–0.77) (seen in Figure 4), with an accuracy of 66.40%, precision of 83.84%, recall of 43.23%, and F1 score of 57.04%. In addition, the AUC of the resting-state EEG-based CNN in depression severity was 0.70 (95%CI: 0.65–0.75) (seen in Figure 5), with an accuracy of 66.93%, precision of 63.49%, recall of 79.69%, and F1 score of 70.67% (seen in Table 3).

2.2.4 Correlation of clinical characteristics and major depression AI score

In the depression group, Spearman's correlation analysis revealed that major depression AI scores were significantly correlated with SAS scores (r = 0.508, p = 0.003) and SDS scores (r = 0.765, p < 0.001) (seen in Figure 6 and Figure7), but the scores were not remarkably correlated with EPQ subscale scores (P, r = 0.011, p = 0.953; E, r = -0.305, p = 0.090; N, r = 0.322, p = 0.072; L, r = 0.208, p = 0.253).

The analysis was also performed between major depression AI scores and SCL-90 subscale scores. It showed that major depression AI scores were significantly correlated with somatization (r = 0.492, p = 0.004), obsessive-compulsive disorder (r = 0.546, p = 0.001), interpersonal sensitivity (r = 0.530, p = 0.002), depression (r = 0.653, p < 0.001), anxiety (r = 0.506, p = 0.003), hostility (r = 0.496, p = 0.004), phobic anxiety (r = 0.485, p = 0.005), paranoid ideation (r = 0.522, p = 0.002), and psychoticism (r = 0.531, p = 0.002).

2.3 Discussion

To the best of our knowledge, there was no previous study that added the EEG screening process in their research. The step was critical considering the vulnerability of the EEG signal (Müller-Putz, 2020). In this study, the AUC of the resting-state EEG-based CNN for differentiating between depression patients and healthy people was 0.74, with an accuracy of 66.40%. The accuracy was within a close approximation to the findings of the previous study (Cai et al., 2018). However, our model got lower AUC and accuracy in depression diagnosis compared to the results of most previous studies (Acharya et al., 2018; Kang et al., 2020; Uyulan et al., 2021). This might be associated with the following reasons: first, some EEG artifacts might be retained, owing to the fewer preprocessing steps (Jiang et al., 2019). The residual EEG artifacts tend to have a great impact on the CNN performance. Second, the amount of data included in this study was relatively small compared to

TABLE 3 Performance of the CNN model in depression diagnosis and its severity.

	AUC	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Depression diagnosis	0.74 (95%CI: 0.70–0.77)	66.40	83.84	43.23	57.04
Depression severity	0.70 (95%CI: 0.65–0.75)	66.93	63.49	79.69	70.67

Note: 95% CI: 95% confidence interval.

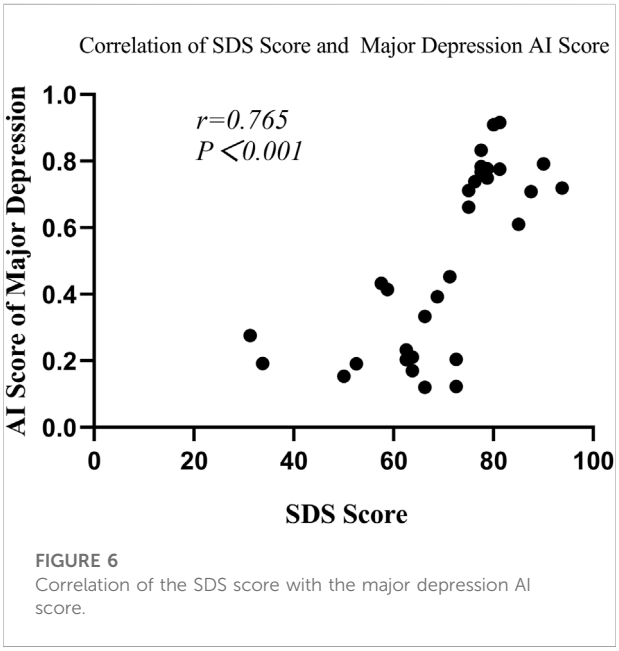


FIGURE 6
Correlation of the SDS score with the major depression AI score.

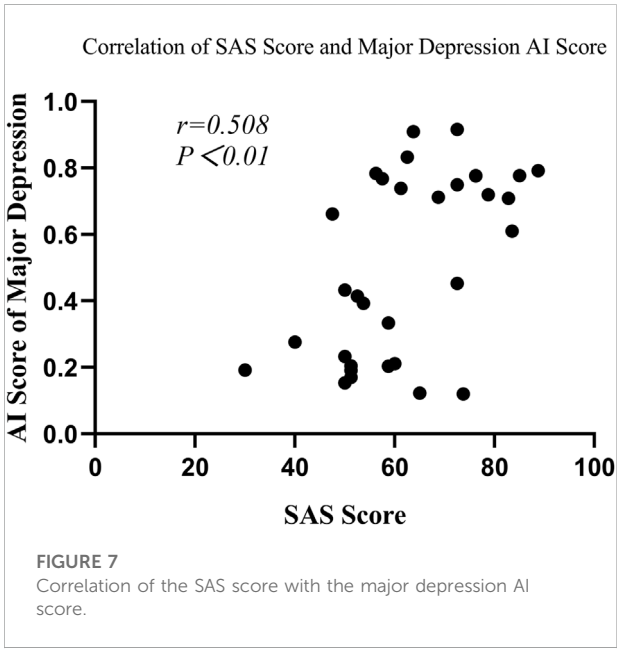


FIGURE 7
Correlation of the SAS score with the major depression AI score.

the previous studies. This also might have influenced the performance of the classification model (Li et al., 2011). Despite that, more steps of EEG preprocessing and feature selection were disadvantageous for assisting clinicians in rapid decision-making. As a result, it suggested that CNN's accuracy in clinical depression diagnosis might have been exaggerated in previous studies.

Furthermore, our study also evaluated depression severity with the model. Similar to the results of previous studies, SDS, SAS, SCL-90 subscales, and N score were significantly higher in the major depression group than those in the non-major depression group. Because people who have co-morbidity with anxiety are more likely to suffer from depression, the enhanced level of anxiety might be related to that of negative emotions (Goldberg and Fawcett, 2012; Choi et al., 2020). In addition, the impairment of cognitive function and strained interpersonal relationships was identified to be the most strongly associated with depression severity (Douglas et al., 2018). Furthermore, this study found that major depression patients had a higher risk of psychotic symptoms (Dubovsky et al., 2021). Major depression with psychotic symptoms tended to have a higher risk of comorbidity and suicide (Gaudiano et al., 2009; Dold et al., 2019), which had a greater impact on the quality of life of patients (Wang et al., 2021). Meanwhile, worse results were anticipated with the treatment of pharmacotherapy and psychotherapy in major depression with psychotic symptoms (Craig et al., 2007; Dold et al., 2019). Therefore, it was particularly important to accurately identify major depression. In this study, the AUC of the resting-state EEG-based CNN for differentiating between non-major depression patients and major depression patients was 0.70 with an accuracy of 66.93%. Although different modeling methods and data processing strategies were used, the AUC and accuracy of our model were close to those of previous studies (Dibeklioglu et al., 2018; Mahato et al., 2020; Kwon and Kim, 2021). It might suggest that the model has better performance and stability in depression grading of severity. Furthermore, we found that major depression AI scores were positively correlated with depression symptoms, which further clarified the aforementioned result.

In this study, major depression AI scores were also positively correlated with anxiety symptoms, somatization,

obsessive-compulsive disorder, interpersonal sensitivity, hostility, phobic anxiety, paranoid ideation, and psychoticism. These results further showed that the aforementioned symptoms were remarkably associated with major depression and indicated that depressed patients with the symptoms were prone to diagnosis with major depression in our model. Meanwhile, it somewhat indicated that the model could be generalized to identify other psychiatric disorders associated with the aforementioned symptoms in the future. In particular, schizophrenia (Shoeibi et al., 2021), bipolar disorder (Li et al., 2021), anxiety disorders (Xing et al., 2019), and obsessive-compulsive disorder (Cohn et al., 2018) should be considered. Unfortunately, we did not find a significant correlation between major depression AI scores and EPQ subscale scores, which were approximately consistent with the clinical results except for neuroticism. It might indicate that personality traits have a limited contribution to identifying depression severity. Meanwhile, it is worth noting that the aforementioned conclusions still should be considered with caution, owing to the restrictions of our results. More analysis will be performed in future research.

In addition, there were still some limitations to this study: first, the lack of questionnaire information on the healthy people and the lack of clinical information on the patients, such as treatment with antidepressants. Second, the groups did not match in age. Third, using a single EEG signal as a data-driven ML model for depression diagnosis lacks clinical value and accuracy compared to existing ML models that include multimodal data. Fourth, we did not build a "diagnostic model of depression" by grouping mild, moderate, and severe image data. Further research can group the three types of data and models separately to obtain a more accurate diagnosis model of depression.

Therefore, further studies with more rigorous experimental design and clinical information are expected. In particular, regional cooperation and multi-center research studies should be encouraged (Geng et al., 2022). The most recent approaches to the diagnosis of depression have focused primarily on graph theory in neuropsychiatry (Aydin et al., 2022; Kilic et al., 2022). Using GT-based network analysis, researchers can estimate global connectivity measures from multichannel EEG recordings. We will pay more attention to network measurement based on graph theory in future research directions. Meanwhile, we would continue to integrate multi-model data such as EEG, fMRI, and DNA methylation data to create more accurate artificial prediction models, eventually providing new strategies for early depression diagnosis and its severity.

3 Conclusion

In this study, our model can accurately identify the depression-specific EEG signal, both in terms of depression diagnosis and its severity identification. Based on this, we conclude that the model could be a useful aid for depression diagnosis and its severity. It would eventually provide new strategies for early depression diagnosis and its severity.

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 studies involving human participants were reviewed and approved by the Research Ethics Board at The First Affiliated Hospital of Nanchang University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

ML: conceptualization, formal analysis, investigation, and writing—original draft. YuL: methodology, investigation, and writing—original draft. YaL: investigation and writing—review and editing. CP: formal analysis and writing—original draft. RY: investigation and writing—review and editing. ZZ: formal analysis. LD and XW: conceptualization, data curation, methodology, investigation, writing—review and editing, and supervision.

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

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., Adeli, H., and Subha, D. P. (2018). Automated EEG-based screening of depression using deep convolutional neural network. *Comput. Methods Programs Biomed.* 161, 103–113. doi:10.1016/j.cmpb.2018.04.012

Ahmadlou, M., Adeli, H., and Adeli, A. (2012). Fractality analysis of frontal brain in major depressive disorder. *Int. J. Psychophysiol.* 85 (2), 206–211. doi:10.1016/j.ijpsycho.2012.05.001

Aydin, S., Cetin, F. H., Uytun, M. C., Babadagi, Z., Gueven, A. S., and Isik, Y. (2022). Comparison of domain specific connectivity metrics for estimation brain network indices in boys with ADHD-C. *Biomed. Signal Process. Control* 76, 103626. doi:10.1016/j.bspc.2022.103626

Bairy, G. M., Bhat, S., Eugene, L. W., Niranjan, U. C., Puthankatti, S. D., Joseph, P. K., et al. (2015). Automated classification of depression electroencephalographic signals using discrete cosine transform and nonlinear dynamicsStudy on Feature Selection Methods for Depression Detection Using Three-Electrode EEG Data. *Interdiscip. Sci.Interdiscip Sci.* 510 (33), 635558–640565. 2018, doi:10.1166/jmihi.2015.1418.Cai10.1007/s12539-018-0292-5

Chen, M., Zhou, L., Ye, L., Lin, G., Pang, Y., Lu, L., et al. (2022). Initial adherence by psychiatric outpatients in a general hospital and relevant personal factors. *BMC Psychiatry* 22 (1), 137. doi:10.1186/s12888-022-03797-3

Choi, K. W., Kim, Y. K., and Jeon, H. J. (2020). Comorbid anxiety and depression: Clinical and conceptual consideration and transdiagnostic treatment. *Adv. Exp. Med. Biol.* 1191, 219–235. doi:10.1007/978-981-32-9705-0_14

Cohn, J. F., Okun, M. S., Jeni, L. A., Ertugrul, I. O., Borton, D., Malone, D., et al. (2018). Automated affect detection in deep brain stimulation for obsessive-compulsive disorder: A pilot study. *Proc. ACM Int. Conf. Multimodal Interact.*, 40–44. doi:10.1145/3242969.3243023

Craig, T. J., Grossman, S., Bromet, E. J., Fochtmann, L. J., and Carlson, G. A. (2007). Medication use patterns and two-year outcome in first-admission patients with major depressive disorder with psychotic features. *Compr. Psychiatry* 48 (6), 497–503. doi:10.1016/j.comppsych.2007.06.005

Čukić, M., López, V., and Pavón, J. (2020). Classification of depression through resting-state electroencephalogram as a novel practice in psychiatry: Review. *J. Med. Internet Res.* 22 (11), e19548. doi:10.2196/19548

Delorme, A., and Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. doi:10.1016/j.jneumeth.2003.10.009

Derogatis, L. R., Lipman, R. S., and Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale-preliminary report. *Psychopharmacol. Bull.* 9 (1), 13–28.

Dibeklioglu, H., Hammal, Z., and Cohn, J. F. (2018). Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE J. Biomed. Health Inf.* 22 (2), 525–536. doi:10.1109/JBHI.2017.2676878

Dimitriadis, S. I. (2021). Reconfiguration of αmplitude driven dominant coupling modes (DoCM) mediated by α-band in adolescents with schizophrenia spectrum disorders. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 108, 110073. doi:10.1016/j.pnpbp.2020.110073

Ding, X., and Yao, J. (2020). Peer education intervention on adolescents' anxiety, depression, and sleep disorder during the COVID-19 pandemic. *Psychiatr. Danub.* 32 (3-4), 527–535. doi:10.24869/psyd.2020.527

Dold, M., Bartova, L., Kautzky, A., Porcelli, S., Montgomery, S., Zohar, J., et al. (2019). Psychotic features in patients with major depressive disorder: A report from the European group for the study of resistant depression. *J. Clin. Psychiatry* 80 (1), 17m12090. doi:10.4088/JCP.17m12090

Douglas, K. M., Gallagher, P., Robinson, L. J., Carter, J. D., McIntosh, V. V., Frampton, C. M., et al. (2018). Prevalence of cognitive impairment in major depression and bipolar disorder. *Bipolar Disord.* 20 (3), 260–274. doi:10.1111/bdi.12602

Duan, L., Duan, H., Qiao, Y., Sha, S., Qi, S., Zhang, X., et al. (2020). Machine learning approaches for MDD detection and emotion decoding using EEG signals. *Front. Hum. Neurosci.* 14, 284. doi:10.3389/fnhum.2020.00284

Dubovsky, S. L., Ghosh, B. M., Serotte, J. C., Cranwell, V., Faust, O., Ang, P. C., et al. (2021). Psychotic depression: Diagnosis, differential diagnosis, and treatment. Psychother PsychosomDepression diagnosis support system based on EEG signal entropies. *J. Mech. Med. Biol.* 9014 (303), 1601450035–1601450177. 2014, doi:10.1142/S0219519414500353

Dunstan, D. A., and Scott, N. (2020). Norms for zung's self-rating anxiety scale. *BMC Psychiatry* 20 (1), 90. doi:10.1186/s12888-019-2427-6

Dunstan, D. A., and Scott, N. (2019). Clarification of the cut-off score for Zung's self-rating depression scale. *BMC Psychiatry* 19 (1), 177. doi:10.1186/s12888-019-2161-0

Gaudiano, B. A., Dalrymple, K. L., and Zimmerman, M. (2009). Prevalence and clinical characteristics of psychotic versus nonpsychotic major depression in a general psychiatric outpatient clinic. *Depress. Anxiety* 26 (1), 54–64. doi:10.1002/da.20470

Geng, H., Chen, J., Chuan-Peng, H., Jin, J., Chan, R. C. K., Li, Y., et al. (2022). Promoting computational psychiatry in China. *Nat. Hum. Behav.* 6, 615–617. doi:10.1038/s41562-022-01328-4

Goldberg, D., and Fawcett, J. (2012). The importance of anxiety in both major depression and bipolar disorder. *Depress. Anxiety* 29 (6), 471–478. doi:10.1002/da.21939

Gong, Y. (1984). Use of the Eysenck personality questionnaire in China. *Pers. Individ. Dif.* 5 (4), 431–438. doi:10.1016/0191-8869(84)90008-4

Grünewald, B. D., Greimel, E., Trinkl, M., Bartling, J., Großheinrich, N., and Schulte-Körne, G. (2018). Resting frontal EEG asymmetry patterns in adolescents with and without major depression. *Biol. Psychol.* 132, 212–216. doi:10.1016/j.biopsycho.2018.01.003

Hosseinifard, B., Moradi, M. H., and Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput. Methods Programs Biomed.* 109 (3), 339–345. doi:10.1016/j.cmpb.2012.10.008

Jiang, X., Bian, G. B., and Tian, Z. (2019). Removal of artifacts from EEG signals: A review. *Sensors* 19 (5), 987. doi:10.3390/s19050987

Kang, M., Kwon, H., Park, J. H., Kang, S., and Lee, Y. (2020). Deep-Asymmetry: Asymmetry matrix image for deep learning method in pre-screening depression. *Sensors (Basel)* 20 (22), 6526. doi:10.3390/s20226526

Khodayari-Rostamabad, A., Reilly, J. P., Hasey, G., Debruin, H., and Maccrimmon, D. (2010). Diagnosis of psychiatric disorders using EEG data and employing a statistical decision model. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 4006–4009. doi:10.1109/IEMBS.2010.5627998

Kilic, B., and Aydin, S. (2022). Classification of contrasting discrete emotional states indicated by EEG based graph theoretical network measures. *Neuroinformatics*. (Online ahead of print). doi:10.1007/s12021-022-09579-2

Kwon, N., and Kim, S. (2021). Depression severity detection using read speech with a divide-and-conquer approach. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2021, 633–637. doi:10.1109/EMBC46164.2021.9629868

Latreille, V., Carrier, J., Gaudet-fex, B., Rodrigues-brazète, J., Panisset, M., Chouinard, S., et al. (2016). Electroencephalographic prodromal markers of dementia across conscious states in Parkinson's disease. *Brain* 139 (pt 4), 1189–1199. doi:10.1093/brain/aww018

Levey, D. F., Polimanti, R., Cheng, Z., Zhou, H., Nuñez, Y. Z., Jain, S., et al. (2019). Genetic associations with suicide attempt severity and genetic overlap with major depression. *Transl. Psychiatry* 9 (1), 22. doi:10.1038/s41398-018-0340-2

Li, D. C., Liu, C. W., and Hu, S. C. (2011). A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artif. Intell. Med.* 52 (1), 45–52. doi:10.1016/j.artmed.2011.02.001

Li, Z., Li, W., Wei, Y., Gui, G., Zhang, R., Liu, H., et al. (2021). Deep learning based automatic diagnosis of first-episode psychosis, bipolar disorder and healthy controls. *Comput. Med. Imaging Graph.* 89, 101882. doi:10.1016/j.compmedimag.2021.101882

Mahato, S., Goyal, N., Ram, D., Paul, S., Shapiro, B., and Rawles, J. (2020). Detection of depression and scaling of severity using six channel EEG data. *J. Med. SystAnn Intern Med.* 44174 (75), 118ITC65–ITC80. doi:10.1007/s10916-020-01573-y.McCarron10.7326/AITC202105180

Morabito, F. C., Campolo, M., Mammone, N., Versaci, M., Franceschetti, S., Tagliavini, F., et al. (2017). Deep learning representation from electroencephalography of early-stage creutzfeldt-jakob disease and features for differentiation from rapidly progressive dementia. *Int. J. Neural Syst.* 27 (2), 1650039. doi:10.1142/S0129065716500398

Moreno-Agostino, D., Wu, Y. T., Daskalopoulou, C., Hasan, M. T., Huisman, M., and Prina, M. (2021). Global trends in the prevalence and incidence of depression:a systematic review and meta-analysis. *J. Affect. Disord.* 281, 235–243. doi:10.1016/j.jad.2020.12.035

Müller-Putz, G. R. (2020). Electroencephalography. *Handb. Clin. Neurol.* 168, 249–262. doi:10.1016/B978-0-444-63934-9.00018-4

Murray, G. K., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., and Wray, N. R. (2021). Could polygenic risk scores Be useful in psychiatry?: A review. *JAMA Psychiatry* 78 (2), 210–219. doi:10.1001/jamapsychiatry.2020.3042

Ortiz, A., Munilla, J., Górriz, J. M., and Ramírez, J. (2016). Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease. *Int. J. Neural Syst.* 26 (7), 1650025. doi:10.1142/S0129065716500258

Puthankattil, S. D., and Joseph, P. K. (2012). Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy. *J. Mech. Med. Biol.* 12 (04), 1240019. doi:10.1142/S0219519412400192

Schönenberg, M., Wiedemann, E., Schneidt, A., Scheeff, J., Logemann, A., Keune, P. M., et al. (2017). Neurofeedback, sham neurofeedback, and cognitive-behavioural group therapy in adults with attention-deficit hyperactivity disorder: A triple-blind, randomised, controlled trial. *Lancet. Psychiatry* 4 (9), 673–684. doi:10.1016/S2215-0366(17)30291-2

Shoeibi, A., Sadeghi, D., Moridian, P., Ghassemi, N., Heras, J., Alizadehsani, R., et al. (2021). Automatic diagnosis of schizophrenia in EEG signals using CNN-lstm models. *Front. Neuroinform.* 15, 777977. doi:10.3389/fninf.2021.777977

Simonato, M., Agoston, D. V., Brooks-Kayal, A., Dulla, C., Fureman, B., Henshall, D. C., et al. (2021). Identification of clinically relevant biomarkers of epileptogenesis - a strategic roadmap. *Nat. Rev. Neurol.* 17 (4), 231–242. doi:10.1038/s41582-021-00461-4

Smith, K. (2014). Mental health: A world of depression. *Nature* 515 (7526), 181. doi:10.1038/515180a

Stewart, J. L., Bismark, A. W., Towers, D. N., Coan, J. A., and Allen, J. J. (2010). Resting frontal EEG asymmetry as an endophenotype for depression risk: Sex-specific patterns of frontal brain asymmetry. *J. Abnorm. Psychol.* 119 (3), 502–512. doi:10.1037/a0019196

Unursaikhan, B., Tanaka, N., Sun, G., Watanabe, S., Yoshii, M., Funahashi, K., et al. (2021). Development of a novel web camera-based contact-free major depressive disorder screening system using autonomic nervous responses induced by a mental task and its clinical application. *Front. Physiol.* 12, 642986. doi:10.3389/fphys.2021.642986

Uyulan, C., Ergüzel, T. T., Unubol, H., Cebi, M., Sayar, G. H., Nezhad Asad, M., et al. (2021). Major depressive disorder classification based on different convolutional neural network models: Deep learning approach. *Clin. EEG Neurosci.* 52 (1), 38–51. doi:10.1177/1550059420916634

Wan, Z. J., Zhang, H., Huang, J. J., Zhou, H. Y., Yang, J., and Zhong, N. (2019). Single-channel EEG-based machine learning method for prescreening major depressive disorder. *Int. J. Info. Tech. Dec. Mak.* 18 (5), 1579–1603. doi:10.1142/S0219622019500342

Wang, M. Q., Wang, R. R., Hao, Y., Xiong, W. F., Han, L., Qiao, D. D., et al. (2021). Clinical characteristics and sociodemographic features of psychotic major depression. *Ann. Gen. Psychiatry* 20 (1), 24. doi:10.1186/s12991-021-00341-7

Wu, W., Zhang, Y., Jiang, J., Lucas, M. V., Fonzo, G. A., Rolle, C. E., et al. (2020). An electroencephalographic signature predicts antidepressant response in major depressionAltered dynamic electroencephalography connectome phase-space features of emotion regulation in social anxiety. *Nat. BiotechnolNeuroimage* 38186 (4), 439338–447349. doi:10.1038/s41587-019-0397-310.1016/j.neuroimage.2018.10.073

Zhang, C., Duan, X. H., Wang, L. Y., Li, Y. L., Yan, B., Hu, G. E., et al. (2021). Dissociable neural representations of adversarially perturbed images in convolutional neural networks and the human BrainUnderstanding the severity of depression: Which symptoms of depression are the best indicators of depression severity? *Compr. PsychiatryCompr Psychiatry* 1587, 67792584–67792588. doi:10.3389/fninf.2021.67792510.1016/j.comppsych.2018.09.006

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 11 October 2022
DOI 10.3389/fmolb.2022.981962

Check for updates

OPEN ACCESS

EDITED BY
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

REVIEWED BY
Adam Brufsky,
University of Pittsburgh Medical Center, United States
Weipeng Zhao,
Tianjin Medical University Cancer Institute and Hospital, China

*CORRESPONDENCE
Aritro Nath,
anath@coh.org
Andrea H. Bild,
abild@coh.org

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 29 June 2022
ACCEPTED 15 September 2022
PUBLISHED 11 October 2022

CITATION
Nath A, Cosgrove PA, Chang JT and Bild AH (2022), Predicting clinical response to everolimus in ER+ breast cancers using machine-learning. *Front. Mol. Biosci.* 9:981962. doi: 10.3389/fmolb.2022.981962

COPYRIGHT
© 2022 Nath, Cosgrove, Chang and Bild. 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.

Predicting clinical response to everolimus in ER+ breast cancers using machine-learning

Aritro Nath[1]*, Patrick A. Cosgrove[1], Jeffrey T. Chang[2] and Andrea H. Bild[1]*

[1]City of Hope Comprehensive Cancer Center, Department of Medical Oncology and Therapeutics, Monrovia, CA, United States, [2]Department of Integrative Biology and Pharmacology, University of Texas Health Science Center at Houston, Houston, TX, United States

Endocrine therapy remains the primary treatment choice for ER+ breast cancers. However, most advanced ER+ breast cancers ultimately develop resistance to endocrine. This acquired resistance to endocrine therapy is often driven by the activation of the PI3K/AKT/mTOR signaling pathway. Everolimus, a drug that targets and inhibits the mTOR complex has been shown to improve clinical outcomes in metastatic ER+ breast cancers. However, there are no biomarkers currently available to guide the use of everolimus in the clinic for progressive patients, where multiple therapeutic options are available. Here, we utilized gene expression signatures from 9 ER+ breast cancer cell lines and 23 patients treated with everolimus to develop and validate an integrative machine learning biomarker of mTOR inhibitor response. Our results show that the machine learning biomarker can successfully distinguish responders from non-responders and can be applied to identify patients that will most likely benefit from everolimus treatment.

KEYWORDS

machine-learning, biomarker, everolimus, estrogen receptor positive breast cancer, prognostic model, random forest, feature selection

Introduction

Breast cancer is now the most commonly diagnosed malignancy and cause of cancer-related death in women worldwide (Houghton and Hankinson, 2021). In the United States, one in eight women will be diagnosed with breast cancer throughout their lifetime (Siegel et al., 2019). At the molecular level, nearly 3 in 4 breast cancers display increased expression of the estrogen receptor (ER+) and do not express the human epidermal growth factor receptor 2 (HER2-). The primary systemic treatment of ER+/HER2- breast cancer is endocrine therapy, which targets the dependency of these tumors on the estrogen signaling pathway for proliferation. These include selective estrogen receptor modulators (SERMs) like tamoxifen, selective estrogen receptor degraders (SERDs) like fulvestrant, and aromatase inhibitors (AIs) like exemestane (Smith and Dowsett, 2003; Patel and Bihani, 2018).

Patients with primary or early-stage ER+/HER2- breast cancer generally have a favorable outlook, with excellent 5-year survival rates on endocrine therapy, even without

the use of adjuvant chemotherapy (Early Breast Cancer Trialists' Collaborative et al., 2012). However, the response rates tend to be lower in patients with metastatic disease, with only 30% of the patients displaying tumor regression on endocrine therapy (Osborne and Schiff, 2011). This outcome has been attributed to primary or acquired endocrine resistance in progressive tumors. Studies have shown that advanced and metastatic ER+ breast cancers can develop endocrine resistance through various mechanisms, such as mutations in the ER-alpha gene, amplification/overexpression of epidermal growth factor and fibroblast growth factor receptor family genes, and activation of downstream signaling *via* the PI3K/AKT/mTOR signaling pathway (Musgrove and Sutherland, 2009; Clarke et al., 2015). Consequently, add-on drugs that target the resistance mechanisms, such as the PI3K inhibitor alpelisib and the mTOR inhibitor everolimus, have shown promising results in clinical trials for advanced ER+ breast cancers. For example, the SOLAR-1 trial reported an improvement of median overall survival in *PIK3CA* mutated cancers from 5.7 months in the fulvestrant group to 11 months in the alpelisib plus fulvestrant group. The BOLERO-2 trial showed significant improvement in progression-free survival in post-menopausal ER+ breast cancers from 2.8 months on exemestane alone to 6.9 months on everolimus plus exemestane (Baselga et al., 2012).

Unlike the PI3K inhibitors, currently, there are no clinically relevant biomarkers available for the selection of everolimus as the treatment for ER+ breast cancers. In the absence of suitable guidelines, this choice is primarily based on patient and caregiver preferences. We have previously shown that effective prognostic and response biomarkers can be developed from the baseline (pre-treatment) transcriptomes of the tumors using systems biology and machine learning (Nath et al., 2019; Nath et al., 2022). In this study, we apply a machine learning framework to develop a novel biomarker model to predict clinical response to everolimus. We adopt a hybrid approach that integrates signatures of treatment response from well-controlled *in vitro* experimentation of cell lines treated with everolimus with empirical signatures derived from the baseline tumor transcriptomes of 23 patients. Using this approach, we develop and validate a predictive model of everolimus response and demonstrate its potential application in identifying candidates for mTOR inhibitor treatment.

Materials and methods

Breast cancer cell line culture and drug treatment

Nine ER+/HER2- breast cancer cell lines were used in this study. CAMA-1, LY2, and MCF7 cell lines were grown and cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco, Cat# 11995073) + 10% heat-inactivated Fetal Bovine Serum (FBS, Sigma-Aldrich, Cat # F4135) + 1x antibiotic-antimycotic (Gibco, Cat# 15240062). T47D, BT-483, ZR-75-1, HCC1428, MDA-MB-134-VI, and MDA-MB-175-VII were grown and cultured in RPMI-1640 (Gibco, Cat# 11875119) + 10% heat-inactivated FBS + 1x antibiotic-antimycotic. Cell lines were authenticated by STR profiling (at *City of Hope Integrative Genomics Core*) and tested negative for mycoplasma contamination using MycoAlert Mycoplasma Detection Kit (Lonza, Cat# LT07-118).

To determine an effective concentration of everolimus and exemestane for everolimus plus exemestane signature each cell line was plated at 1,000 cell/well in a 384-well flat bottom TC-treated plate (Corning, Cat# 3764) and allowed to adhere at 37°C humidified incubator + 5% CO_2. After 24hrs post-plating, cells were incubated with a dose-response of everolimus or exemestane or 0.2% DMSO control for 4 days (40 µL total volume). Viability was assessed as a measure of total ATP using the CellTiter-Glo assay (Promega, Cat# G7573) according to manufacturer instructions. See Supplementary Table S1 and Supplementary Figure S1.

Each cell line was plated at 250,000 cells/well in 2 ml of the respective culture media on a 6-well tissue culture treated plate (Costar, Cat# 3506) and allowed to adhere at 37°C humidified incubator + 5% CO_2. After 24 h post-plating, the cells were treated with either 0.2% DMSO (control) or a combination of 0.5 nM everolimus + 25 µM exemestane (Selleck Chemicals, Cat# S1120, S1196 respectively) in their respective culture media. Following treatment, the cells were incubated for 6 h at 37°C + 5% CO_2 in a humidified incubator.

Cell lines RNA extraction, cDNA synthesis, library preparation, and sequencing

After 6 h of treatment, the plated cells were rinsed one time with ice-cold 1x PBS (Gibco, Cat# 10010049) followed by collection *via* cell scrapping in ice-cold 1x PBS. Collected cells were stored frozen overnight at -80°C in RNAlater (Invitrogen, Cat# AM7023). Frozen cells were thawed at 4°C, washed in 1x PBS, and RNA was isolated using the AllPrep DNA/RNA Mini Kit (Qiagen, Cat# 80204) according to the manufacturer's instructions. Sequencing libraries were prepared and ran at Fulgent Genetics (Temple City, CA) using NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs, Cat# E7760L) and sequenced on Illumina NovaSeq 6000 with S4 flow cell (2 × 150 cycles) with 20 M PE reads per sample.

Cell lines RNA-seq data preprocessing

Raw sequencing read files (fastq) were pre-processed using the Bioinformatics ExperT System (BETSY) (Chen and Chang, 2017). Sequencing quality was assessed using FastQC and

adapter trimming was performed using trimmomatic (0.33) (Bolger et al., 2014). Sequences were aligned using STAR (2.7.6a) (Dobin et al., 2013), followed by counts estimation using HTSeq (Anders et al., 2015) and estimation of gene expression levels using RSEM (1.3.1) (Li and Dewey, 2011). Transcript per million (TPM) values from RSEM were log2(x+1) transformed and filtered to remove genes with the lowest variance (25th percentile) and lowest expression (30th percentile).

Developing signature for *in vitro* everolimus response using bayesian binary regression

Filtered TPM matrix for the nine cell lines was used to train a supervised Bayesian binary regression model based on the method developed by West et al. (West et al., 2001) and implemented in the GenePattern module SIGNATURE (Chang et al., 2011). Expression values were quantile normalized and a set of 100 features (genes) were obtained that were correlated with the treatment status (DMSO vs. everolimus plus exemestane). A prediction model based on Bayesian regression that used the two metagenes (principal component of the signature gene matrix) with Monte-Carlo simulations was used to obtain classification accuracy in leave-one-out cross-validation (LOOCV) analysis. Based on the successful classification of the cell lines using this model, the selected features were used for further analysis.

Patient microarray data preprocessing

Gene expression data from patients in a neoadjuvant everolimus trial (Sabine et al., 2010) were obtained from NCBI GEO accession GSE119262. We used expression data from the pre-treatment tumors to train and validated the model. The tumor samples were profiled using Illumina HumanRef-8 v2 Expression BeadChips and quantile normalized using BeadArray (Sabine et al., 2010). We aggregated the expression matrix by first averaging data from multiple probes at the gene level and then averaging the expression levels of replicates. The log-transformed expression levels were standardized such that each gene had a mean = 0 and standard deviation = 1 across the samples.

Integrative machine-learning framework for response prediction

We implemented a LOOCV framework using the caret package for R (Kuhn, 2021) to combine the *in vitro* signature genes with genes selected from the clinical dataset to develop an integrative biomarker. In each iteration of the cross-validation, we first selected a set of relevant features using Fast Correlation Based Filter for Feature Selection implemented (FCBF) using the FCBF package for R (Lubiana and Nakaya, 2021). We then obtained an integrative signature by combining the *in vitro* signature and the FCBF selected features and used this set of genes as predictors in a random forest model, with the patient response as the outcome variable. This was performed using the randomForest R package (Liaw and Wiener, 2002). An internal cross-validation was performed within each iteration to tune the mtry hyperparameter. All analyses were performed in R version 4.1.0 (R Core Team, 2021).

Functional enrichment analyses

Pathway enrichment analyses were performed using the g:Profiler2 package for R (Kolberg et al., 2020). Genes in the *in vitro* signature were split into two lists (up or down in everolimus treated cells) and analyzed for enrichment of GO:BP, KEGG, and REACTOME pathway terms. Functional enrichment was performed for the over-representation of genes using the hypergeometric test and adjusted for multiple comparisons. Enrichment plots and tables were created using the g:Profiler2 package, with color coding in the tables showing the level of evidence associated with the terms. A dark blue color indicated weaker evidence whereas an orange color indicated strong, experimentally derived evidence for the term.

Results

An integrative machine learning framework

We developed an integrated biomarker development approach that harnessed evidence from controlled *in vitro* experiments with ER+/HER2- breast cancer cell lines treated with everolimus and combined this with data from a neoadjuvant clinical trial of ER+ breast cancer patients treated with everolimus. The outline for our approach is shown in Figure 1. First, we cultured nine cancer cell lines, including MCF7, T47D, CAMA1, ZR-75-1, HCC1428, MDA-MB-134, BT483, LY2 and MDA-MB-175 in either 0.2% DMSO (control) or a combination of 0.5 nM everolimus + 25 μM exemestane. The treatment concentration for the experiment was determined based on the dose-response curves of the nine cell lines (Supplementary Figure S1; Supplementary Table S1). After 6 h of treatment, total RNA was extracted from each pair of untreated and treated cell lines and sequenced at a target read depth of 20 M reads. We then filtered the pre-processed gene expression (RSEM) from each cell line to retain the most informative genes by removing low

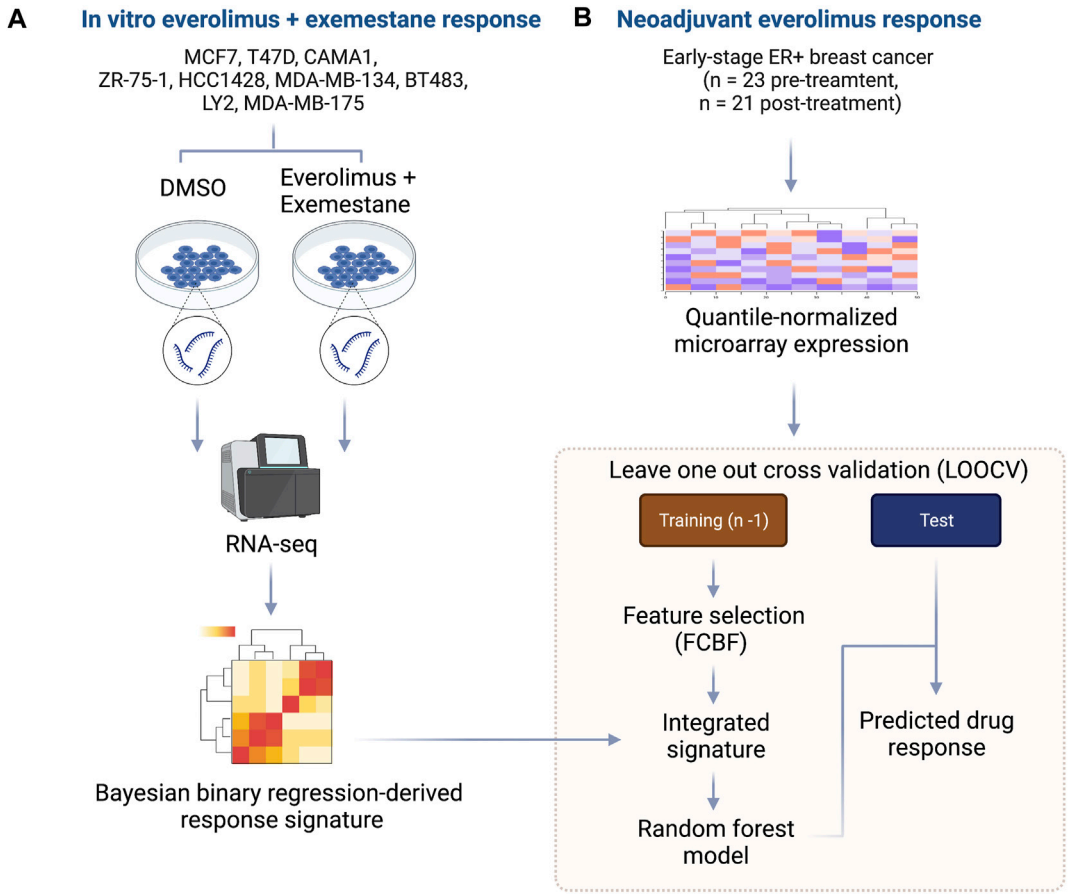


FIGURE 1
Outline of integrative approach for mTOR inhibitor biomarker development. **(A)** The *in vitro* signature was developed using 9 ER+ breast cancer cell lines. Each cell line was treated with either DMSO or everolimus plus exemestane. Total RNA was extracted, and cDNA libraries were prepared for RNA-seq. The raw transcripts were pre-processed, followed by quantile normalization and feature selection using the Bayesian binary regression framework. **(B)** Microarray data from early-stage ER+ breast cancer patients treated with neoadjuvant everolimus were pre-processed and analyzed within a leave-one-out cross-validation (LOOCV) framework. Each iteration of the LOOCV generated a list of features correlated with response. These features were integrated with the response signature derived from cell lines to obtain an integrated signature. The integrated signature was then used as a set of predictors in random forest model to predict the response in the test sample.

expression and low variance genes. The expression levels were quantile normalized, followed by feature selection, and fitting a Bayesian binary regression model with treatment status as the outcome (Figure 1A). Concurrently, we obtained gene expression data from pre-treatment biopsies of 23 ER+ breast cancers that received neoadjuvant everolimus for about 2 weeks (Sabine et al., 2010). This trial reported clinical response as a change in Ki67 staining percentage at the end of 11–14 days of treatment, with patients showing more than a 10% decrease in Ki67 staining classified as responders. We then implemented a LOOCV framework that used two sets of features: 1. A set of signature genes from used in the Bayesian binary regression model of the cell lines treated with everolimus and exemestane and 2. A set of features that were selected using FCBF. This integrated set of features was used to train a random forest classifier within each fold of the LOOCV (Figure 1B).

Transcriptomic signature of *in vitro* everolimus response

We created an *in vitro* everolimus response signature using RNA-seq profile of the nine ER+ breast cancer cells, with the treatment status (DMSO vs. everolimus plus exemestane) as the outcome variable. Starting with a matrix of filtered gene expression data across cell lines, we first defined the signature set by selecting genes using Pearson correlation that best differentiated the cell lines based on treatment status (Figure 2A). A Bayesian binary regression model was then fit

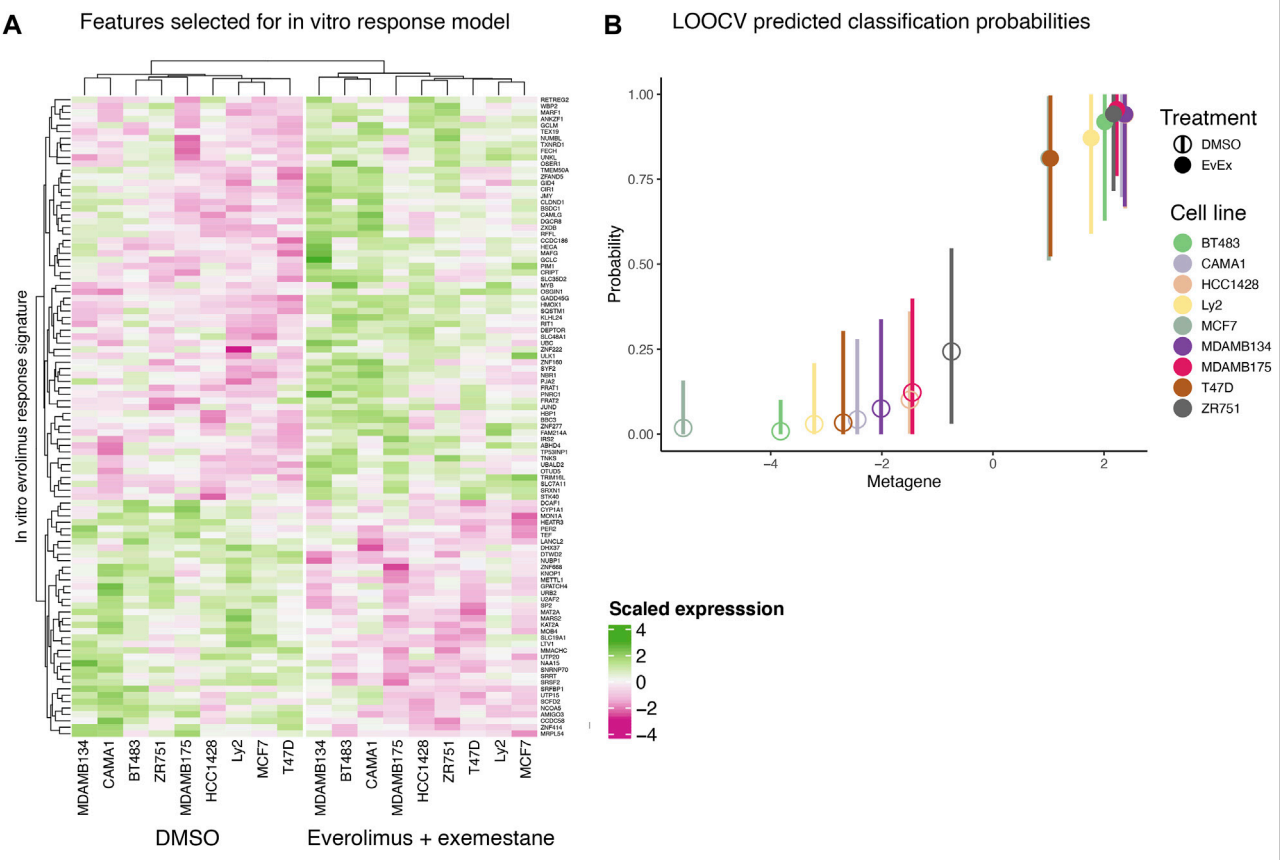


FIGURE 2
In vitro everolimus response signature and validation. **(A)** Heatmap of genes selected by the binary regression model to classify and predict everolimus response in 9 ER+/HER2- breast cancer cell lines. The genes in the signature are listed in rows and the columns indicate cell lines. Both rows and columns are shown as hierarchical clusters, with the columns split into two clades, resulting in clustering by treatment status (DMSO and everolimus plus exemestane). **(B)** LOOCV analysis of the 9 breast cancer cell lines. Hollow circles indicate cell lines treated with DMSO while the solid circles indicate cell lines treated with everolimus plus exemestane. The X-axis indicates metagene score, calculated from the principal component of the genes in the signature. Y-axis indicates predicted probability of response, with a value closer to 1 indicating response. The vertical bars indicate 95% confidence interval of the prediction probability.

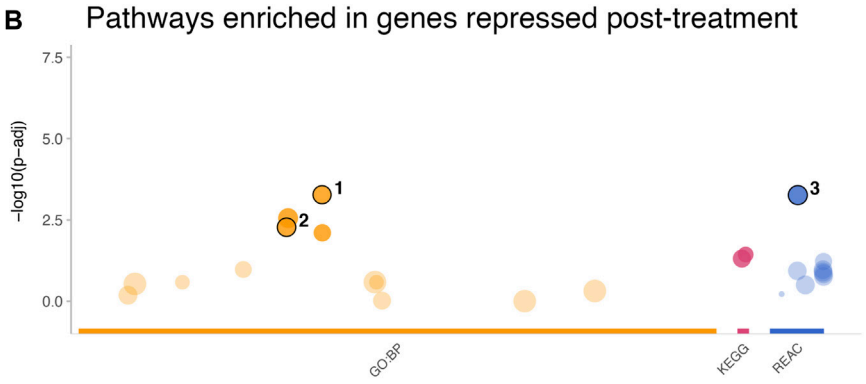
on the first two principal components of the signature gene expression matrix to classify the cells. This model was sampled using a Markov chain Monte Carlo algorithm to obtain posterior probabilities and 95% confidence intervals (Figure 2B). A probability closer to 1 indicated that the signature genes were active in cells treated with everolimus. As shown in Figure 2B, the signature could clearly distinguish cell lines based on treatment status.

Further examination of the signature genes revealed key biological processes and pathways activated or inactivated in the cell lines post treatment. Enrichment analysis for GO:BP, KEGG and REACTOME terms in the genes expressed at higher levels in cells treated with everolimus revealed activation of pathways related to cell death and apoptosis (Figure 3A; Supplementary Table S2). For example, some of the key enriched pathways included response to oxidative stress, regulation of apoptosis, ferroptosis and pexophagy, which are well-known consequences of mTOR inhibition *in vitro*. On the other hand, genes that were downregulated were enriched in terms associated with translation and cell proliferation (Figure 3B; Supplementary Table S3). Again, this agreed with the expectation that mTOR inhibition would lead to reduced protein turnover and proliferation rates.

Predicting clinical response using integrative model

A clinical study of ER+ breast cancers evaluated everolimus response by measuring percentage Ki67 staining change over the course of treatment of 2 weeks (Sabine et al., 2010). This clinical trial reported response data from 23 pre-treatment biopsies and 21 post-treatment biopsies. We used the pre-treatment gene

id	source	term_id	term_name	term_size	p_value
1	GO:BP	GO:0006979	response to oxidative stress	420	3.2e−03
2	GO:BP	GO:0042981	regulation of apoptotic process	1419	6.2e−03
3	KEGG	KEGG:04216	Ferroptosis	39	3.2e−04
4	REAC	REAC:R−HSA−9664873	Pexophagy	11	1.4e−03



id	source	term_id	term_name	term_size	p_value
1	GO:BP	GO:0042254	ribosome biogenesis	304	5.3e−04
2	GO:BP	GO:0034470	ncRNA processing	429	5.3e−03
3	REAC	REAC:R−HSA−8953854	Metabolism of RNA	661	5.4e−04

FIGURE 3
Enrichment analysis of in vitro gene signature. **(A,B)** The dotplots show significance of the enrichment terms from GO:BP, KEGG and REACTOME signatures, with Y-axis showing −log10 of the FDR-adjusted *p*-value from the enrichment test. Key significant terms enriched in genes that were **(A)** expressed at higher levels in the treated cells or **(B)** expressed at higher cells in the untreated cells are annotated in the table below. The term size indicated number of genes in the original signature, while the color code indicates strength of evidence associated with the term. Terms supported by experimental evidence are shown in orange.

expression data to develop a biomarker that can predict response to everolimus treatment. This analysis was performed within an LOOCV framework, where each iteration of the cross validation involved selecting relevant features associated with treatment response in the training split, integrating the selected features with the *in vitro* gene signature, and training and validation of a random forest model. We used FCBF algorithm to select the features associated with treatment response. This algorithm utilized symmetrical uncertainty, an information theory derived concept that selected genes with high correlation with the outcome but low correlation with other variables. The genes selected with FCBF were integrated with the *in vitro* signature to train and evaluate the random forest model. LOOCV analysis showed that the predicted probabilities of response based on pre-treatment gene

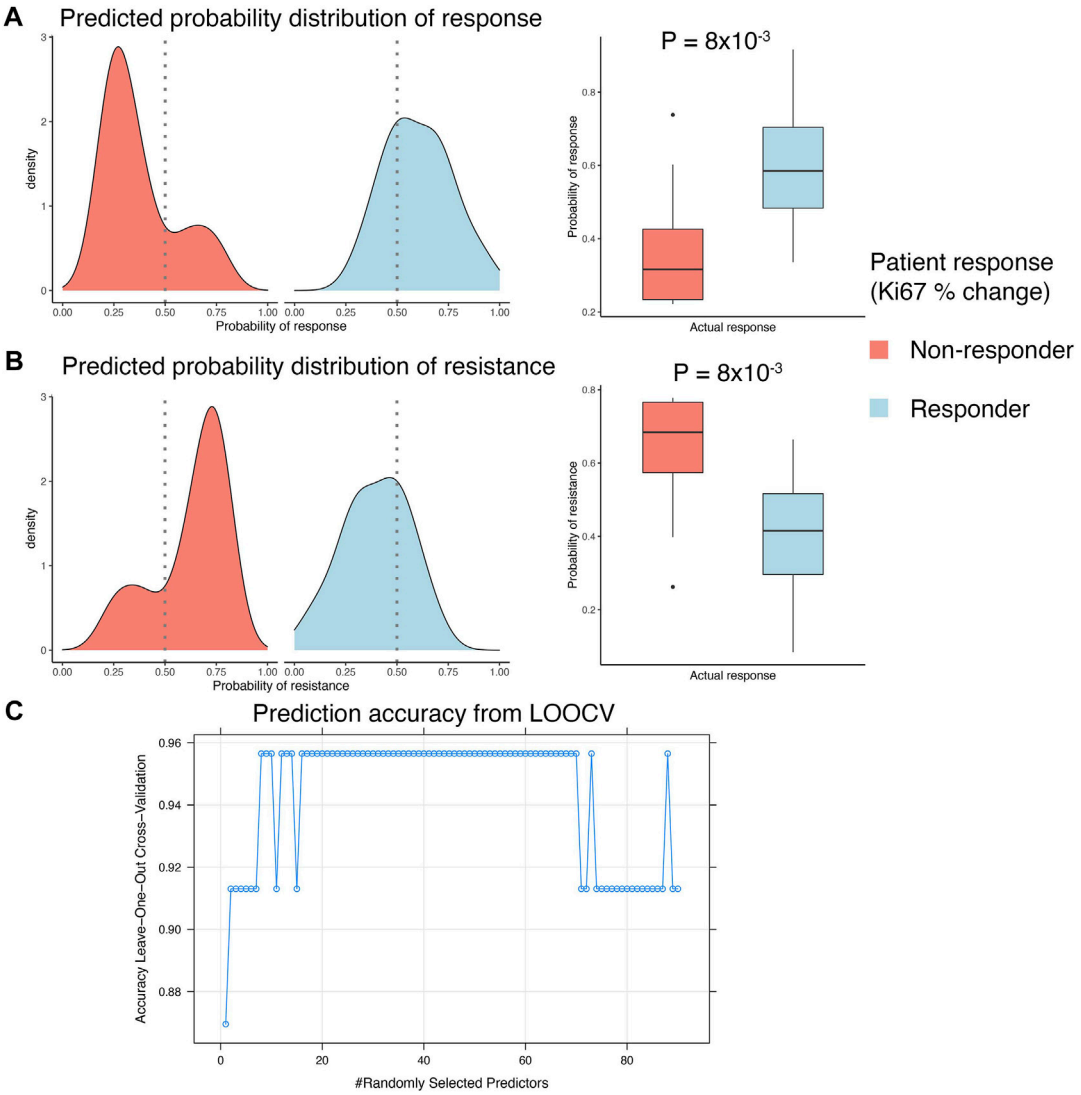


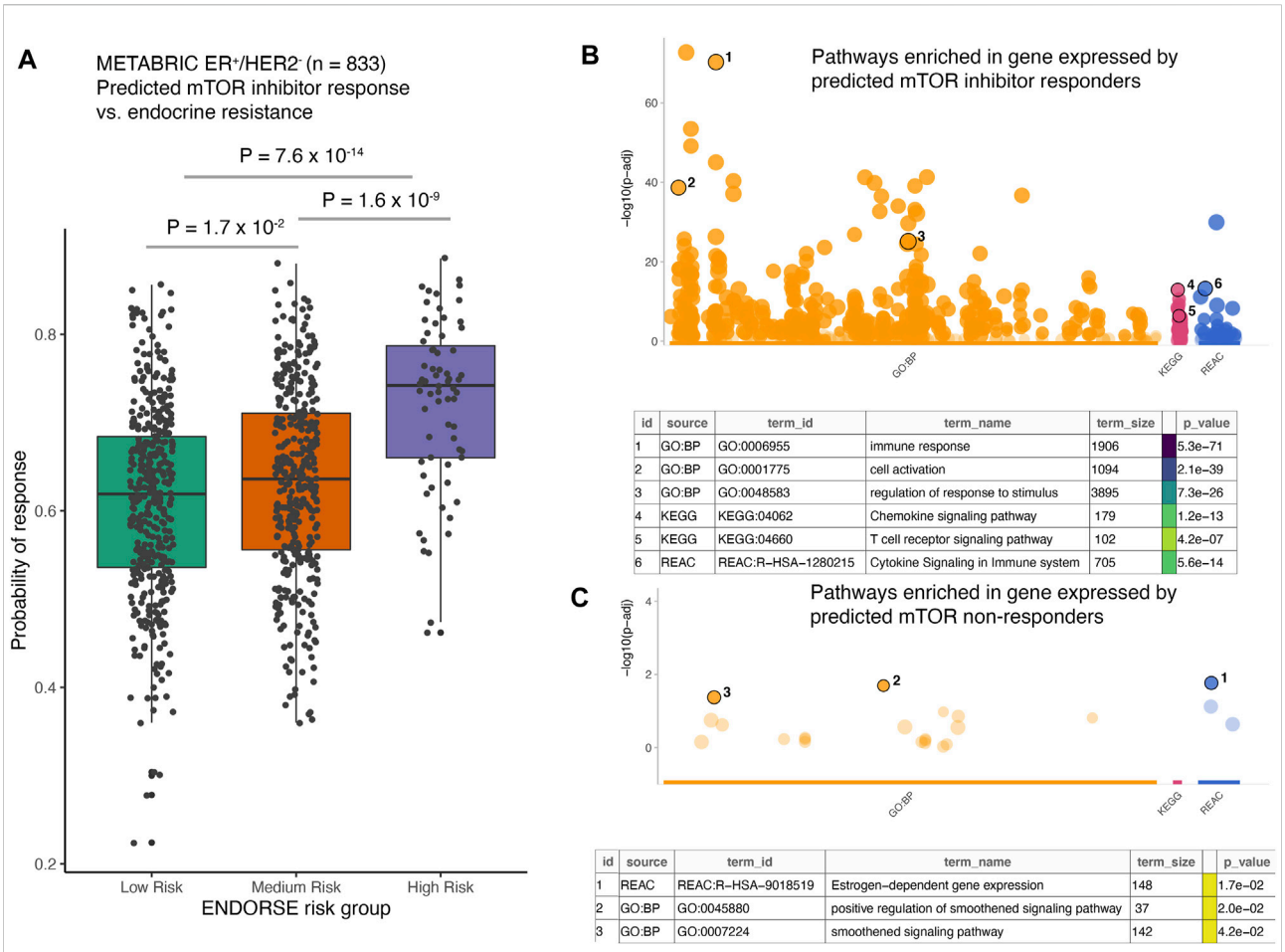
FIGURE 4
Developing integrated model with patient response data. **(A,B)** LOOCV analysis of the GSE119262 datasets comparing the prediction probabilities of **(A)** response or **(B)** non-response (resistance) calculated using the integrated model combining the *in vitro* response signature with the FCBF-selected features. The density plots on the left show distribution of the prediction probabilities in samples grouped by actual clinical response, with red indicating patients that were clinical non-responders (<10% decrease in Ki67% staining after 2–4 weeks) and blue indicating patients that were clinical responders (>10% decrease in Ki67% staining after 2–4 weeks). The boxplots on the right show statistical comparison of the prediction probabilities between patients grouped by actual clinical response. **(C)** Dot plot showing trends in change of accuracy of the random forest model in LOOCV analysis with varying values of the mtry hyperparameter. An accuracy of ~0.95 was achieved with mtry = 8 and used to construct the final prediction model.

expression data agreed with the actual clinical outcomes, as reported by the clinical trial (Figure 4A). Similarly, the predicted probabilities of non-response agreed with the reported clinical response (Figure 4B). Overall, the random forest model fit on the complete clinical dataset of 23 pre-treatment tumor samples was highly accurate, with consistently high accuracies (>0.9) achieved in the LOOCV analyses for tuning the hyperparameters of the random forest model fit on the full dataset (Figure 4C).

Identifying potential candidates for everolimus treatment

Given the high prediction accuracy of the model in LOOCV analyses, we applied the random forest model trained on the full clinical dataset of 23 pre-treatment samples to predict mTOR inhibitor response in the METABRIC cohort of ER+/HER2- tumors (Curtis et al., 2012). This cohort included 833 breast cancer patients that

10.3389/fmolb.2022.981962



id	source	term_id	term_name	term_size	p_value
1	GO:BP	GO:0006955	immune response	1906	5.3e−71
2	GO:BP	GO:0001775	cell activation	1094	2.1e−39
3	GO:BP	GO:0048583	regulation of response to stimulus	3895	7.3e−26
4	KEGG	KEGG:04062	Chemokine signaling pathway	179	1.2e−13
5	KEGG	KEGG:04660	T cell receptor signaling pathway	102	4.2e−07
6	REAC	REAC:R−HSA−1280215	Cytokine Signaling in Immune system	705	5.6e−14

id	source	term_id	term_name	term_size	p_value
1	REAC	REAC:R−HSA−9018519	Estrogen−dependent gene expression	148	1.7e−02
2	GO:BP	GO:0045880	positive regulation of smoothened signaling pathway	37	2.0e−02
3	GO:BP	GO:0007224	smoothened signaling pathway	142	4.2e−02

FIGURE 5
Application of the mTOR inhibitor model on external dataset. **(A)** The boxplots show predicted probability of mTOR inhibitor response in METABRIC ER+/HER2- patients (*n* = 833). The patients were classified as endocrine sensitive (low-risk), endocrine intermediate (medium-risk) and endocrine resistant (high-risk) using the ENDORSE model. The adjusted *p*-values annotated above the boxplots show pairwise comparisons obtained from Tukey's HSD test applied to a one-way ANOVA model **(B,C)**. The dotplots show significance of the enrichment terms from GO: BP, KEGG and REACTOME signatures, with Y-axis showing -log10 of the FDR-adjusted *p*-value from the enrichment test. Key significant terms enriched in genes that were **(B)** expressed at higher levels in the tumors predicted to be responsive to mTOR inhibitor treatment or **(C)** expressed at higher levels in the tumors non-responsive to mTOR inhibitor treatment are annotated in the table below. The term size indicated number of genes in the original signature, while the color code indicates strength of evidence associated with the term. Terms supported by weak evidence are shown in blue while experimentally derived signatures are shown in orange.

had received endocrine therapy and were either alive at study completion or died due to the disease. We had previously developed a biomarker model to predict patients with high risk of death on endocrine therapy in this cohort. This model, called ENDORSE, could successfully stratify patients based on predicted endocrine resistance. We had also noted that the patients with ENDORSE risks showed activation of the mTOR signaling pathway. Therefore, we compared the predicted probabilities of mTOR inhibitor response with the ENDORSE classes in the METABRIC cohort. Here, we found that the predicted mTOR response were significantly higher in medium and high-risk groups than the low-risk groups (Figure 5A). Moreover, a large proportion of the high-risk tumors (>40%) showed a high probability of mTOR inhibitor response (>0.75) compared to medium-risk (15%) or low-risk tumors (10%). We further investigated the biological signals enriched in the tumors with a high probability of mTOR inhibitor response (>0.75). Interestingly, we found an overwhelming majority of the biological processes and signaling pathways at the top of the list of significant terms to be associated with immune signaling and communication (Figure 5B; Supplementary Table S4). In contrast, the signatures enriched in non-responders were associated with estrogen signaling or

smoothened signaling pathways (Figure 5C; Supplementary Table S5). These suggested the tumors predicted to be non-responsive to mTOR inhibitors were still largely dependent on estrogen signaling or bypassed mTOR signaling *via* the smoothened pathway for growth and proliferation.

Discussion

The use of mTOR inhibitors such as everolimus has shown promising results in improving outcomes of ER+/HER2- breast cancer patients(Ellard et al., 2009; Baselga et al., 2012; Andre et al., 2014; Piccart et al., 2014). However, there are several treatment options available for progressive and advanced ER+ breast cancers, which include drugs that target the PI3K and mTOR signaling pathways. While the presence of activating mutations on the *PIK3CA* gene guide the use of PI3K inhibitors such as alpelisib (Narayan et al., 2021), the decision to use mTOR inhibitors like everolimus is completely based on patient and provider choice. Therefore, our goal was to systematically develop a new biomarker that may be useful in predicting clinical outcomes for mTOR inhibitors. We have previously developed a prognostic model for endocrine resistance in breast cancer patients using the tumor baseline transcriptomic data (Nath et al., 2022). Here, we extended our approaches to develop a novel machine learning biomarker for everolimus response.

To date, only one clinical trial with any mTOR inhibitor has reported genomic information from breast cancer patients before and after treatment (Sabine et al., 2010). This trial was of limited size of only 23 pre-treatment samples, which made it challenging to train and develop an effective biomarker model. We addressed this issue by first identifying which genes are expressed in response to everolimus in a well-controlled *in vitro* environment. Such approaches have been extensively used to generate gene expression signatures by directly manipulating the expression of genes or using chemical perturbations *in vitro*. For example, the curated and oncogenic signature collection in the molecular signatures database contains over 3000 such signatures generated using genetic or chemical perturbations (Liberzon et al., 2011) that are frequently used in prognostic and drug response signatures (Sonachalam et al., 2012; Tan et al., 2019; Kong et al., 2020; Zeng et al., 2022).

As the *in vitro* environment is less affected by the inter-sample variances typically observed in animal and patient-derived data, our approach allowed us to pick genes with high confidence that show a significant change upon everolimus treatment and are likely good candidate features for a machine learning biomarker model (Figure 2). These genes were sensible and associated with expected biological phenomenon (Figure 3). We then implemented an approach that leveraged pre-treatment tumor transcriptomes and clinical outcomes data from the 23 patients combined with the *in vitro* signature. This integrated model was highly accurate in predicting clinical everolimus response in the LOOCV analyses of the patient data (Figure 4).

We further applied the biomarker to predict mTOR inhibitor response in an independent cohort of ER+/HER2- breast cancer patients from the METABRIC study (Curtis et al., 2012). We obtained gene expression and overall survival data from 833 patients. These patients had received only endocrine therapy and were classified using a prognostic model that predicted risk of death on endocrine therapy (Nath et al., 2022). We had previously found that the METABRIC patients with high risk of death on endocrine therapy showed elevated pathway activity of PI3K/AKT/mTOR signaling pathway (Nath et al., 2022). By applying our biomarker, we found that indeed a vast proportion of the predicted mTOR responsive patients were those in the high risk group (Figure 5). Activation of mTOR signaling is a well-documented phenomenon associated with endocrine resistance and poor prognosis of ER+ breast cancer patients (Ciruelos Gil, 2014; Paplomata and O'Regan, 2014; Dong et al., 2021; Nunnery and Mayer, 2020). Thus, our novel biomarker could be useful in identifying the patients that are most likely to benefit from mTOR inhibitor treatment.

Another interesting aspect of our study were the biological signatures and pathways activated *in vitro* upon everolimus treatment and the ones enriched in patients predicted to be responsive. The *in vitro* signature largely showed enrichment of expected biological pathways, including cellular oxidative stress (Piao et al., 2014), autophagy (Crazzolara et al., 2009) and apoptosis (Tai et al., 2017). In comparison, the patient data showed a large proportion of immune activation pathways as the most significant signatures. mTOR signaling is well-known to play an important role in directing adaptive immune response by receiving microenviromental signals and activating T-cells and dendritic cells (Delgoffe and Powell, 2009). In the tumor microenvironment, mTOR signaling regulates the activity of macrophages and T-cells through inflammatory factors like IL-10, TGF-beta, and membrane bound CTLA-4 and PD-1 (Kim et al., 2017). This has been linked with a shift in balance from an anti-tumor to a pro-tumor immune microenvironment by reducing the proportion of anti-tumor $CD8^+$ T-cells and increasing the proportion of Treg and tumor-associated M2 macrophages (Kim et al., 2017; Mafi et al., 2021). Thus, the enrichment of these immune activation-related terms in the patient data captures a known effect of elevated mTOR pathway activity and supports the biomarker-driven classification of the patients as likely responders to mTOR inhibition. We also observed an enrichment of smoothened receptor pathway signatures in the mTOR-resistant tumor. As a major component of the hedgehog signaling pathway, both the canonical and non-canonical activation of the smoothened pathway has been linked with stem-cell like traits, invasiveness and metastatic progression of breast cancers (Jeng et al., 2020). Consequently, multiple interventions

targeting the hedgehog and smoothened signaling pathway are currently being evaluated in breast cancer (Bhateja et al., 2019).

Key limitations of this study are lack of a large-scale training dataset and an additional independent validation dataset for the biomarker model. The clinical training data used in the study consisted of only 23 pre-treatment samples, with a large number of potential predictive features. We attempted to mitigate this challenge by systematically reducing the number of predictive features using the *in vitro* signature and selecting a limited number of empirical features from the clinical datasets for model construction. Furthermore, we performed the model development and validation in a LOOCV framework, where the empirical features from the clinical dataset were only picked from the training split. The model was then applied to predict outcome in the left-out test sample. This approach helped in diminishing problems associated with overfitting models to the data and overestimating model accuracies. Nevertheless, clinical translation of the biomarker will benefit greatly from additional validation and refinement using prospective biopsies or through retrospective analyses of banked samples.

Given that patients progressing on endocrine therapy have multiple treatment options, including aromatase inhibitors, chemotherapy, PI3K inhibition or mTOR blockage, development of biomarkers to guide therapy selection of these patients can help ensure they are treated with the most effective drug regimen. This study uses both experimental and patient-based data to develop a biomarker for response to everolimus, and to understand the signaling underlying inhibition of mTOR signaling in ER+ breast cancer.

Data availability statement

The neoadjuvant everolimus clinical trial dataset used in this study can be found in the NCBI GEO repository under the accession GSE119262 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119262). The METABRIC dataset can be obtained from the cBioPortal for cancer genomics (https://www.cbioportal.org/study/summary?id=brca_metabric). Scripts associated with this manuscript are available at https://github.com/aritronath/Everolimus_MLB. The GenePattern BinReg module, results, and the RNA-seq input data for the nine ER+ breast cancer cell lines analyzed in this study are also available in this GitHub repository under "releases". Genes in the final integrative signature are provided in Supplementary Table 6.

Author contributions

AN: Conceptualization; Data curation; Software; Formal analysis; Validation; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. PC: Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. JC: Data curation; Software; Formal analysis; Validation; Investigation. AB: Conceptualization; Resources; Supervision; Funding acquisition; Investigation; Methodology; Project administration; Writing—review and editing.

Funding

Research reported in this publication was supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award number U54CA209978 awarded to AB, JC, and a pilot grant awarded through U54CA209978 to AN.

Acknowledgments

The authors would like to acknowledge the work provided by the Leadership and Staff of the CoH Center for Informatics most notably Research Informatics, and the utilization of the POSEIDON data exploration, visualization, and analysis platform including the Honest Broker process. Figure 1 was created with BioRender.com.

Conflict of interest

AN and AB are listed co-inventors on a patent application (filed, pending) covering the mTOR inhibitor response gene signature described in this manuscript.

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/fmolb.2022.981962/full#supplementary-material

References

Anders, S., Pyl, P. T., and Huber, W. (2015). Htseq--A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi:10.1093/bioinformatics/btu638

Andre, F., O'regan, R., Ozguroglu, M., Toi, M., Xu, B., Jerusalem, G., et al. (2014). Everolimus for women with trastuzumab-resistant, her2-positive, advanced breast cancer (Bolero-3): A randomised, double-blind, placebo-controlled phase 3 trial. *Lancet. Oncol.* 15, 580–591. doi:10.1016/S1470-2045(14)70138-X

Baselga, J., Campone, M., Piccart, M., Burris, H. A., 3rd, Rugo, H. S., Sahmoud, T., et al. (2012). Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N. Engl. J. Med.* 366, 520–529. doi:10.1056/NEJMoa1109653

Bhateja, P., Cherian, M., Majumder, S., and Ramaswamy, B. (2019). The hedgehog signaling pathway: A viable target in breast cancer? *Cancers* 11, E1126. doi:10.3390/cancers11081126

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Chang, J. T., Gatza, M. L., Lucas, J. E., Barry, W. T., Vaughn, P., and Nevins, J. R. (2011). Signature: A workbench for gene expression signature analysis. *Bmc Bioinforma.* 12, 443. doi:10.1186/1471-2105-12-443

Chen, X., and Chang, J. T. (2017). Planning Bioinformatics workflows using an expert system. *Bioinformatics* 33, 1210–1215. doi:10.1093/bioinformatics/btw817

Ciruelos Gil, E. M. (2014). Targeting the pi3k/akt/mtor pathway in estrogen receptor-positive breast cancer. *Cancer Treat. Rev.* 40, 862–871. doi:10.1016/j.ctrv.2014.03.004

Clarke, R., Tyson, J. J., and Dixon, J. M. (2015). Endocrine resistance in breast cancer--an overview and update. *Mol. Cell. Endocrinol.* 418, 220–234. doi:10.1016/j.mce.2015.09.035

Crazzolara, R., Bradstock, K. F., and Bendall, L. J. (2009). Rad001 (everolimus) induces autophagy in acute lymphoblastic leukemia. *Autophagy* 5, 727–728. doi:10.4161/auto.5.5.8507

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., et al. (2012). The genomic and transcriptomic architecture of 2, 000 breast tumours reveals novel subgroups. *Nature* 486, 346–352. doi:10.1038/nature10983

Delgoffe, G. M., and Powell, J. D. (2009). Mtor: Taking cues from the immune microenvironment. *Immunology* 127, 459–465. doi:10.1111/j.1365-2567.2009.03125.x

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal rna-seq aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Dong, C., Wu, J., Chen, Y., Nie, J., and Chen, C. (2021). Activation of pi3k/akt/mtor pathway causes drug resistance in breast cancer. *Front. Pharmacol.* 12, 628690. doi:10.3389/fphar.2021.628690

Early Breast Cancer Trialists' Collaborative, G., Peto, R., Davies, C., Godwin, J., Gray, R., Pan, H. C., et al. (2012). Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100, 000 women in 123 randomised trials. *Lancet* 379, 432–444. doi:10.1016/S0140-6736(11)61625-5

Ellard, S. L., Clemons, M., Gelmon, K. A., Norris, B., Kennecke, H., Chia, S., et al. (2009). Randomized phase ii study comparing two schedules of everolimus in patients with recurrent/metastatic breast cancer: Ncic clinical trials group Ind.163. *J. Clin. Oncol.* 27, 4536–4541. doi:10.1200/JCO.2008.21.3033

Houghton, S. C., and Hankinson, S. E. (2021). Cancer progress and priorities: Breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 30, 822–844. doi:10.1158/1055-9965.EPI-20-1193

Jeng, K.-S., Sheen, I. S., Leu, C.-M., Tseng, P.-H., and Chang, C.-F. (2020). The role of smoothened in cancer. *Int. J. Mol. Sci.* 21, E6863. doi:10.3390/ijms21186863

Kim, L. C., Cook, R. S., and Chen, J. (2017). Mtorc1 and Mtorc2 in cancer and the tumor microenvironment. *Oncogene* 36, 2191–2201. doi:10.1038/onc.2016.363

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). Gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion. *Toolset G:Profiler.* F1000res, 9.

Kong, J., Lee, H., Kim, D., Han, S. K., Ha, D., Shin, K., et al. (2020). Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. *Nat. Commun.* 11, 5485. doi:10.1038/s41467-020-19313-8

Kuhn, M. (2021). *Caret: Classification and regression training.* Version 6.0-88 Ed. R Package.

Li, B., and Dewey, C. N. (2011). Rsem: Accurate transcript quantification from rna-seq data with or without A reference genome. *Bmc Bioinforma.* 12, 323. doi:10.1186/1471-2105-12-323

Liaw, A., and Wiener, M. (2002). Classification and regression by randomforest. *R News* 2 (3), 18–22.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (msigdb) 3.0. *Bioinformatics* 27, 1739–1740. doi:10.1093/bioinformatics/btr260

Lubiana, T., and Nakaya, H. (2021). *Fcbf: Fast correlation based filter for feature selection.* R Package.Version 2.0.0.

Mafi, S., Mansoori, B., Taeb, S., Sadeghi, H., Abbasi, R., Cho, W. C., et al. (2021). Mtor-mediated regulation of immune responses in cancer and tumor microenvironment. *Front. Immunol.* 12, 774103. doi:10.3389/fimmu.2021.774103

Musgrove, E. A., and Sutherland, R. L. (2009). Biological determinants of endocrine resistance in breast cancer. *Nat. Rev. Cancer* 9, 631–643. doi:10.1038/nrc2713

Narayan, P., Prowell, T. M., Gao, J. J., Fernandes, L. L., Li, E., Jiang, X., et al. (2021). Fda approval summary: Alpelisib plus fulvestrant for patients with hr-positive, her2-negative, pik3ca-mutated, advanced or metastatic breast cancer. *Clin. Cancer Res.* 27, 1842–1849. doi:10.1158/1078-0432.CCR-20-3652

Nath, A., Cohen, A. L., and Bild, A. H. (2022). Endorse: A prognostic model for endocrine therapy in estrogen-receptor-positive breast cancers. *Mol. Syst. Biol.* 18, E10558. doi:10.15252/msb.202110558

Nath, A., Lau, E. Y. T., Lee, A. M., Geeleher, P., Cho, W. C. S., and Huang, R. S. (2019). Discovering long noncoding rna predictors of anticancer drug sensitivity beyond protein-coding genes. *Proc. Natl. Acad. Sci. U. S. A.* 116, 22020–22029. doi:10.1073/pnas.1909998116

Nunnery, S. E., and Mayer, I. A. (2020). Targeting the pi3k/akt/mtor pathway in hormone-positive breast cancer. *Drugs* 80, 1685–1697. doi:10.1007/s40265-020-01394-w

Osborne, C. K., and Schiff, R. (2011). Mechanisms of endocrine resistance in breast cancer. *Annu. Rev. Med.* 62, 233–247. doi:10.1146/annurev-med-070909-182917

Paplomata, E., and O'regan, R. (2014). The pi3k/akt/mtor pathway in breast cancer: Targets, trials and biomarkers. *Ther. Adv. Med. Oncol.* 6, 154–166. doi:10.1177/1758834014530023

Patel, H. K., and Bihani, T. (2018). Selective estrogen receptor modulators (serms) and selective estrogen receptor degraders (serds) in cancer treatment. *Pharmacol. Ther.* 186, 1–24. doi:10.1016/j.pharmthera.2017.12.012

Piao, S. G., Lim, S. W., Doh, K. C., Jin, L., Heo, S. B., Zheng, Y. F., et al. (2014). Combined treatment of tacrolimus and everolimus increases oxidative stress by pharmacological interactions. *Transplantation* 98, 22–28. doi:10.1097/TP.0000000000000146

Piccart, M., Hortobagyi, G. N., Campone, M., Pritchard, K. I., Lebrun, F., Ito, Y., et al. (2014). Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: Overall survival results from BOLERO-2†. *Ann. Oncol.* 25, 2357–2362. doi:10.1093/annonc/mdu456

R Core Team (2021). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation For Statistical Computing.

Sabine, V. S., Sims, A. H., Macaskill, E. J., Renshaw, L., Thomas, J. S., Dixon, J. M., et al. (2010). Gene expression profiling of response to mtor inhibitor everolimus in pre-operatively treated post-menopausal women with oestrogen receptor-positive breast cancer. *Breast Cancer Res. Treat.* 122, 419–428. doi:10.1007/s10549-010-0928-6

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *Ca. Cancer J. Clin.* 69, 7–34. doi:10.3322/caac.21551

Smith, I. E., and Dowsett, M. (2003). Aromatase inhibitors in breast cancer. *N. Engl. J. Med.* 348, 2431–2442. doi:10.1056/NEJMra023246

Sonachalam, M., Shen, J., Huang, H., and Wu, X. (2012). Systems biology approach to identify gene network signatures for colorectal cancer. *Front. Genet.* 3, 80. doi:10.3389/fgene.2012.00080

Tai, S., Xu, L., Xu, M., Zhang, L., Zhang, Y., Zhang, K., et al. (2017). Combination of arsenic trioxide and everolimus (Rad001) synergistically induces both autophagy and apoptosis in prostate cancer cells. *Oncotarget* 8, 11206–11218. doi:10.18632/oncotarget.14493

Tan, T. Z., Ye, J., Yee, C. V., Lim, D., Ngoi, N. Y. L., Tan, D. S. P., et al. (2019). Analysis of gene expression signatures identifies prognostic and functionally distinct ovarian clear cell carcinoma subtypes. *Ebiomedicine* 50, 203–210. doi:10.1016/j.ebiom.2019.11.017

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., et al. (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11462–11467. doi:10.1073/pnas.201162998

Zeng, A. G. X., Bansal, S., Jin, L., Mitchell, A., Chen, W. C., Abbas, H. A., et al. (2022). A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat. Med.* 28, 1212–1223. doi:10.1038/s41591-022-01819-x

frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 14 October 2022
DOI 10.3389/fphys.2022.951368

Check for updates

OPEN ACCESS

EDITED BY
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

REVIEWED BY
Chi-Wen Lung,
Asia University, Taiwan
Ashley Weaver,
Wake Forest University, United States
Diana Madrid Fuentes,
Wake Forest University, United States in collaboration with reviewer AW

*CORRESPONDENCE
Sokratis Makrogiannis,
smakrogiannis@desu.edu

SPECIALTY SECTION
This article was submitted to Computational Physiology and Medicine,
a section of the journal
Frontiers in Physiology

RECEIVED 23 May 2022
ACCEPTED 26 September 2022
PUBLISHED 14 October 2022

CITATION
Makrogiannis S, Okorie A, Di Iorio A, Bandinelli S and Ferrucci L (2022), Multi-atlas segmentation and quantification of muscle, bone and subcutaneous adipose tissue in the lower leg using peripheral quantitative computed tomography.
Front. Physiol. 13:951368.
doi: 10.3389/fphys.2022.951368

COPYRIGHT
© 2022 Makrogiannis, Okorie, Di Iorio, Bandinelli and Ferrucci. 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.

Multi-atlas segmentation and quantification of muscle, bone and subcutaneous adipose tissue in the lower leg using peripheral quantitative computed tomography

Sokratis Makrogiannis[1]*, Azubuike Okorie[1], Angelo Di Iorio[2], Stefania Bandinelli[3] and Luigi Ferrucci[4]

[1]Math Imaging and Visual Computing Lab, Division of Physics, Engineering, Mathematics and Computer Science, Delaware State University, Dover, DE, United States, [2]Antalgic Mini-invasive and Rehab-Outpatients Unit, Department of Innovative Technologies in Medicine & Dentistry, University "G.d'Annunzio", Chieti-Pescara, Italy, [3]Geriatric Unit, Local Health Unit Tuscany Centre, Florence, Italy, [4]National Institute on Aging, National Institutes of Health, Baltimore, MD, United States

Accurate and reproducible tissue identification is essential for understanding structural and functional changes that may occur naturally with aging, or because of a chronic disease, or in response to intervention therapies. Peripheral quantitative computed tomography (pQCT) is regularly employed for body composition studies, especially for the structural and material properties of the bone. Furthermore, pQCT acquisition requires low radiation dose and the scanner is compact and portable. However, pQCT scans have limited spatial resolution and moderate SNR. pQCT image quality is frequently degraded by involuntary subject movement during image acquisition. These limitations may often compromise the accuracy of tissue quantification, and emphasize the need for automated and robust quantification methods. We propose a tissue identification and quantification methodology that addresses image quality limitations and artifacts, with increased interest in subject movement. We introduce a multi-atlas image segmentation (MAIS) framework for semantic segmentation of hard and soft tissues in pQCT scans at multiple levels of the lower leg. We describe the stages of statistical atlas generation, deformable registration and multi-tissue classifier fusion. We evaluated the performance of our methodology using multiple deformable registration approaches against reference tissue masks. We also evaluated the performance of conventional model-based segmentation against the same reference data to facilitate comparisons. We studied the effect of subject movement on tissue segmentation quality. We also applied the top performing method to a larger out-of-sample dataset and report the quantification results. The results show that multi-atlas image segmentation with diffeomorphic deformation and probabilistic label fusion produces very good quality over all tissues, even for scans with significant quality degradation. The application of our technique to the larger dataset reveals trends of age-related body composition changes that are consistent with the literature.

Because of its robustness to subject motion artifacts, our MAIS methodology enables analysis of larger number of scans than conventional state-of-the-art methods. Automated analysis of both soft and hard tissues in pQCT is another contribution of this work.

KEYWORDS

tissue segmentation, tissue quantification, multi-atlas techniques, subject movement, clinical application, pQCT

1 Introduction

Accurate segmentation of tissues using medical imaging is key for the quantification of changes in the structure and composition of tissues, which may result from diseases, aging, and other risk factors related to the tissue(s) in question (Cordova et al., 2014; Töpfer et al., 2015; Owen et al., 2019; Rodrigues and Pinheiro 2019; Wong and Manske 2020). Recent advances of artificial intelligence (AI) in the field of medical imaging have also drawn the interest of researchers in the application of computer techniques in the area of bone and muscle imaging (Burns et al., 2020). In clinical studies, segmentation of bone, muscle, and adipose tissue can be used for computing objective measures and descriptors of body composition and for exploring the causes and effects of differences of these descriptors between subject groups (Lauretani et al., 2008; Makrogiannis et al., 2018).

In the past 2 decades, peripheral quantitative computed tomography (pQCT) and high resolution peripheral quantitative computed tomography (HR-pQCT) have emerged as essential technologies for segmentation and quantification of bone, muscle and adipose tissue properties at the diaphyseal regions of the limbs. Segmentation of hard and soft tissues in pQCT and HR-pQCT imaging has been used to assess the effects of type-2 diabetes mellitus (T2DM) (Starr et al., 2018), osteoporosis (Simon et al., 2022) and osteoarthritis (Chen et al., 2018), to establish measures for characterizing sex-, ethnic-, site-, and age-related outcomes (Gabel et al., 2018), to study the effect of exercise on the muscle and fat cross-sectional areas (Rowe et al., 2019), and in studies of aging and age-related diseases (Chow et al., 2022; Liu et al., 2022). A challenge in pQCT-based segmentation is subject movement and the associated motion artifacts. Subject movement, subtle or obvious, occurs frequently in standard pQCT and HR-pQCT scans (Wong 2016). It may degrade the image quality and affect the assessment of bone and muscle properties (Pialat et al., 2012; Chan et al., 2018). pQCT motion has been assessed by visual inspection followed by a pass or fail decision. Usual criteria are the presence of discontinuities and streaks at the cortical bone and changes in intensity of trabecular bone. Motion streaks originating from the cortical bone extend into the muscle. Quantitative evaluation methods have been proposed for pQCT (Blew et al., 2014) and HR-pQCT (Pauchard et al., 2011; Sode et al., 2011). Thresholding (Blew et al., 2014) and watershed segmentation techniques (Wong 2016) have been employed to identify and assess motion streaks in the muscle. Motion artifact correction and automatic segmentation are desirable.

However, to the best of our knowledge, there is no previous report in the literature on pQCT segmentation techniques explicitly addressing subject movement and the limited contrast-to-noise ratio that are characteristic of this modality. In this work we propose to address this gap by developing a multi-atlas image segmentation (MAIS) framework (Rohlfing et al., 2001, 2005; Shen and Hammer, 2002; Langerak et al., 2010; Sotiras et al., 2013; Iglesias and Sabuncu 2015) for identification of soft and hard tissues in pQCT scans of the lower leg. The MAIS framework includes the stages of statistical atlas generation, linear and non-linear registration, and label fusion for tissue segmentation. In these stages we use pQCT images of the lower leg at 4%, 38%, and 66% of the tibial length. We validated segmentation performance against manual reference masks using the Dice Similarity Coefficient (DSC). We evaluated the performance of multiple atlas-based tissue segmentation techniques and an established model-based tissue segmentation technique. We expect that the segmentation performance of multi-atlas based methods is largely unaffected by motion. We have also applied our framework to a larger out-of-sample dataset and reported our results on age-related tissue composition changes. Figure 1 summarizes the main components of the proposed framework and related experiments.

2 Our methodology

2.1 Atlas-based tissue segmentation

We formulate the problem of atlas-based segmentation next. Given a subject S, an atlas A and its atlas label map SA, we aim to produce the segmentation of S by warping the atlas to the spatial domain of the subject. This stage is called image registration, and is followed by pixel-wise assignment of tissue labels from the warped segmentation atlas to the subject (Figure 2). Since a single atlas is used for segmentation, we refer to this method as *single atlas image segmentation (SAIS)*.

In multi-atlas image segmentation (MAIS), multiple atlases A_i, $i = 1, \ldots, N$ and corresponding segmented atlases SA_i, $i = 1$,

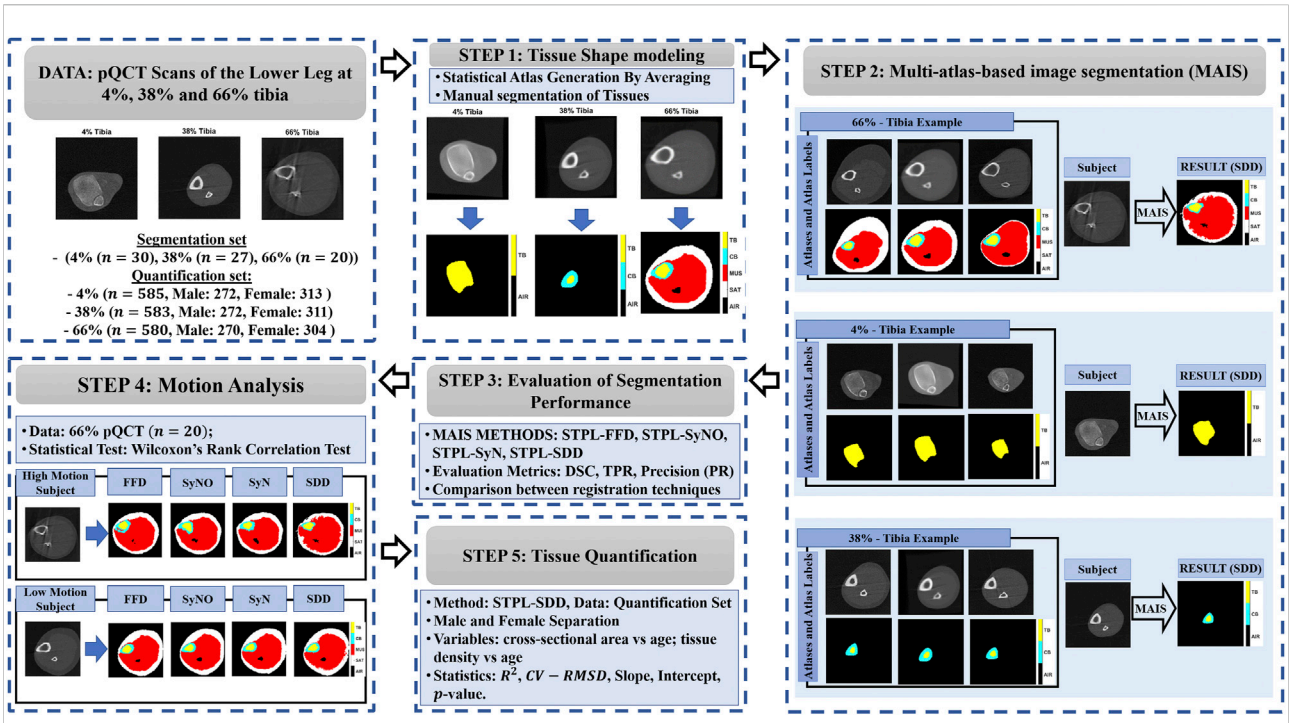


FIGURE 1
Main components of the proposed framework. STAPLE (STPL), free-form deformation (FFD), symmetric diffeomorphic demons (SDD), symmetric normalization (SyN), symmetric normalization - only (SyNO), dice similarity coefficient (DSC), true positive rate (TPR), precision (PR) squared-Spearman's correlation coefficient (R^2), and coefficient of variation - root mean squared difference (CV-RMSD).

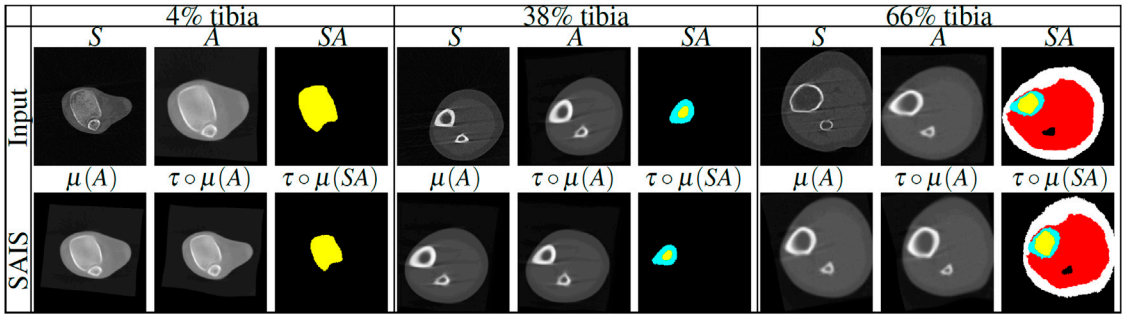


FIGURE 2
Illustration of single atlas based image segmentation (SAIS) stages of lower leg scans at 4%, 38% and 66% tibia. First row: subject (*S*), statistical atlas (*A*) and segmented atlas label map (*SA*) for each tibial length that are used as inputs for segmentation. Second row: linear registration, nonlinear registration and label propagation output. The tissue labels are color-coded as follows: trabecular bone (yellow), cortical bone (cyan), muscle (red), and SAT (white). μ denotes linear transformation, $\tau \circ \mu$ denotes the composition of linear (μ) with nonlinear (τ) transformations.

. . . , *N* are used to produce a segmentation of the subject *S*. The main stages of a multi-atlas-based segmentation algorithm are *registration, label propagation* and *label fusion*. Figure 3 shows the main stages of MAIS. In the following subsections, we elaborate on our method based on these stages. We will use *A* when referring to each atlas A_i, $i \in \{1, \ldots, N\}$.

2.2 Image registration

Image registration is a key stage of atlas-based segmentation as described in the previous section. The goal of registration is to align the spatial domain of a subject with that of an atlas. In other words, we wish to find the optimal

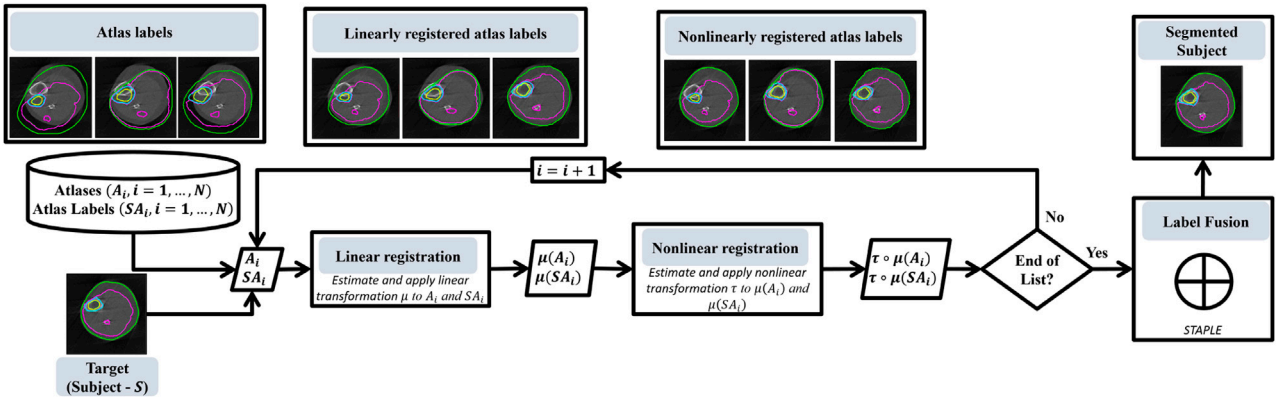


FIGURE 3
Flowchart that shows the main stages of our multi-atlas based segmentation methodology.

TABLE 1 Tissues to quantify in each anatomical site.

Anatomical site	Tissues to quantify
4% tibia	Trabecular Bone (TB)
38% tibia	Cortical Bone (CB), Trabecular Bone
66% tibia	Cortical Bone, Trabecular Bone, Subcutaneous Adipose Tissue (SAT), Muscle

deformation τ^* in the set $\{\tau,\ \tau : (A,\ \Omega_A) \rightarrow (S,\ \Omega_S)\}$ of all transformations from the spatial domain of the source image (or atlas) $(A,\ \Omega_A)$ to the subject space $(S,\ \Omega_S)$, that minimizes the following energy functional (Sotiras et al., 2013; Vercauteren et al., 2007),

$$\mathcal{E}(\tau) \coloneqq \lambda_\zeta \zeta(S, A \circ \tau) + \lambda_\rho \rho(\tau), \tag{1}$$

where ζ is the similarity term, ρ is the regularization term, λ_ζ and λ_ρ are the weights of the similarity and regularization terms, respectively. Therefore,

$$\tau^* = \arg\min_{\tau} \mathcal{E}(\tau). \tag{2}$$

To ensure accurate registration we applied both linear and nonlinear registrations. Linear registration is used to capture the rigid displacement of the subject while nonlinear registration is used to capture the local deformation of the anatomical structures of the subject. Anatomical structures are the tissue types, i.e., the trabecular bone, cortical bone, muscle, and subcutaneous adipose tissue as shown in Table 1. Figure 2 shows examples of rigid displacements of atlas A denoted by $\mu(A)$, and local deformations $\tau \circ \mu$ produced by nonlinear registration $\tau \circ \mu$. Our techniques for linear and nonlinear registration are described below.

2.2.1 Linear registration

Given a subject S and for each atlas A, we estimate the parameters of a linear transformation μ from atlas space (A, Ω_A) to the subject space (S, Ω_S) that defines the rigid motion between the atlas and the subject. Our linear transformation is modeled using affine transformations. We utilize the Mattes' mutual information similarity measure (Mattes et al., 2003) given in Eq. 3,

$$\zeta(S, A \circ \mu) = -\sum_{\iota} \sum_{\kappa} p(\iota, \kappa|\mu) \log \frac{p(\iota, \kappa|\mu)}{p_A(\iota|\mu) p_S(\kappa)} \tag{3}$$

and a regular step gradient descent optimizer, to estimate the parameters of the affine transformation μ,where p, is the joint probability distribution of subject and the atlas, p_A and p_S are the marginal probability distributions of the atlas and the subject respectively, $\iota = 1, \ldots, n_A$ and $\kappa = 1, \ldots, n_S$ are the indices of the histogram bins for the source and target image. The image $\mu(A)$ of the atlas A lives in the subject space and approximates the rigid motion between the subject and the atlas (Figure 2).

2.2.2 Nonlinear registration

In this stage, the goal is to correct the local deformations between the subject and the atlas. Hence we seek the parameters of deformation $\tau : (\mu(A), \Omega_S) \rightarrow (S, \Omega_S)$ from $\mu(A) \subset S$ onto S. We introduce the use of three nonlinear deformable registration techniques, free-form deformation (FFD), symmetric diffeomorphic demons (SDD), and symmetric image normalization (SyN, SyNO) for our multi-atlas-based registration. We discuss the nonlinear methods for our multi-atlas-based image segmentation below.

2.2.2.1 Free-form deformations

Free-form deformation was originally proposed by Rueckert et al., 1999 and was applied to automated registration of breast

MRI scans. It uses a spline interpolation kernel to compute the deformation values between the control points that produces a locally controlled, globally smooth transformation.

Given a 2D spatial domain $\Omega = \{\mathbf{x} = (x, y) | 0 \le x < X, 0 \le y < Y\}$ of an image, let Φ denote an $n_x \times n_y$ mesh of control points $\phi_{i,j} = (i\delta, j\delta)$ with uniform spacing δ. Rueckert et al., 1999 proposed a method that seeks the optimal FFD (displacement field τ) written as the 2-D tensor product of 1-D cubic B-Splines

$$\tau(\mathbf{x}) = \sum_{l=0}^{2} \sum_{m=0}^{2} B_l(u) B_m(v) \phi_{i+l, j+m} \tag{4}$$

that optimizes the energy functional in Eq. 1. In Eq. 4, $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$, and B_l represents the lth basis function of the B-spline. The regularization term ρ is given by the bending energy of a thin-plate of metal, which controls the smoothness of the transformation, defined by

$$\rho(\tau) = \frac{1}{|\Omega|} \iint_{\Omega} \left(\frac{\partial \tau}{\partial x} + \frac{\partial \tau}{\partial y} \right)^2 d\Omega, \tag{5}$$

where $|\Omega|$ is the cardinality of the image spatial domain. We employed Matte's mutual information (Mattes et al., 2003) defined in Eq. 3 as the similarity measure ζ. We used $\lambda_\zeta = -1$ and $\lambda_\rho = 0.01$ in the energy function of Eq. 1. The optimization process is based on updating control points *via* the gradient of the cost function. We employed Limited-memory Broyden–Fletcher–Goldfarb–Shannon optimization to find the energy minimum of Eq. 1. We embedded this method in a hierarchical multi-scale structure to be able to capture a wide range of deformations. This structure contains a coarse and fine scale for optimization. At the coarse scale, the optimizer can capture extensive deformations, but the solution may have limited precision. So we use the coarse solution to initialize the optimization at the fine scale and find a more precise deformation field.

2.2.2.2 Symmetric diffeomorphic demons

We utilize a variant of the Demons algorithm that is optimized in the log domain as proposed in Vercauteren et al., 2008. This is a variational method that seeks to minimize the following energy functional:

$$\mathcal{E}(c, \tau) = \frac{1}{\lambda_\zeta^2} \frac{1}{2|\Omega_\mathbf{p}|} \sum_{\mathbf{x} \in \Omega_\mathbf{x}} |S(\mathbf{x}) - A(\tau(\mathbf{x}))|^2 + \frac{1}{\lambda_h^2} \|\log(\tau^{-1} \circ c)\|^2 + \frac{1}{\lambda_\rho^2} \|\nabla \log(\tau)\|^2 \tag{6}$$

where the variable c was introduced to approximate the error in the correspondence between image pixels, λ_h accounts for spatial uncertainty on the correspondences, and $\Omega_\mathbf{p}$ is the region of overlap between S and $A \circ \tau$.

In the update step, and under the assumption that the current transformation τ is expressible as an exponential of smooth vector fields $\mathbf{v}$, i.e., $\tau = \exp(\mathbf{v})$, the Baker-Campbell Hausdorff (BCH) approximations are used to seek a smooth velocity field $Z(\mathbf{v}, \varepsilon\mathbf{u})$, such that $\exp(Z(\mathbf{v}, \varepsilon\mathbf{u}) \approx \tau \circ \exp(\varepsilon\mathbf{u})$, where ε is a weight parameter. Then $\mathbf{u}$ is given by

$$\mathbf{u}(\mathbf{x}) = -\left(\frac{(S(\mathbf{x}) - A \circ \tau(\mathbf{x}))}{\left(\|J^P\|^2 + \left(\lambda_\zeta^2(\mathbf{x}) / \lambda_h^2 \right) \right)} \right) J^{\mathbf{x}^\top}, \tag{7}$$

and J is the Jacobian matrix. In the log-domain, the inverse of a spatial transform τ^{-1}, parametrized by $\tau = \exp(\mathbf{v})$, can be obtained efficiently by backward computation $\tau^{-1} = \exp(-\mathbf{v})$. A symmetric transformation can be obtained from a nonsymmetric one by making the global energy symmetric, i.e.,

$$\tau_{opt} = \arg\min_{\tau} \left(\mathcal{E}_{sym} := \mathcal{E}(S, A \circ \tau) + \mathcal{E}(A, S \circ \tau^{-1}) \right). \tag{8}$$

The minimization of the energy functional in Eq. 8 can be formulated as a constrained equation using two diffeomorphisms

$$\left[\tau_{opt}, \tau_{opt}^{-1} \right] = \arg\min_{[\tau, \tau^{-1}]} \mathcal{E}_{sym}. \tag{9}$$

2.2.2.3 Symmetric image normalization

This method uses diffeomorphisms as the transformation model (Avants et al., 2008). SyN performs the normalization by minimizing the energy functional defined in Eq. 10. SyN searches for a symmetric diffeomorphic spatiotemporal mapping, $\tau \in Diff_0$: = {the space of diffeomorphic mappings with homogeneous boundary conditions} that minimizes the energy functional in the optimization problem defined in Eq. 10.

$$\mathcal{E}_{sym}(S, A) = \inf_{\tau_1} \inf_{\tau_2} \int_{t=0}^{\frac{1}{2}} \{ \|v_1(\tau_1(\mathbf{x}, t), t)\|_L^2 + \|v_2(\tau_2(\mathbf{x}, t), t)\|_L^2 \} dt + \int_\Omega \zeta(S(\tau_2(0.5), A(\tau_1(0.5)) d\Omega$$
$$\text{subject to each } \tau \in Diff_0 \text{ the solution of } : d\tau(\mathbf{x}, t)/dt = v_i(\tau_i(\mathbf{x}, t), t)$$
$$\text{with } \tau_i(\mathbf{x}, 0) = \mathbf{I}, \text{and } \tau_i^{-1}(\tau_i) = \mathbf{I}, \ \tau_i(\tau_i^{-1}) = \mathbf{I} \tag{10}$$

The first integral in Eq. 10 corresponds to the regularization term that is induced by a functional norm $\|\cdot\|_L$ through a linear differential operator $L = a\nabla^2 + b\mathbf{I}$, with constants a and b, and $\mathbf{I}$ is the identity mapping. The second integral corresponds to the similarity term between the reference and input image, where ζ is Mattes' mutual information defined in Eq. 3, Ω is the common spatial domain of the images, $v(\mathbf{x}, t)$ is the velocity field, and t is the time. The optimization process performs gradient descent to update the deformation field and a fixed point method calculates the inverse transformation. The velocity fields v_i are computed iteratively, and they update the deformation τ_i, $i = 1, 2$. The deformable registration stage is preceded by rigid and affine transformation steps to address global misalignments as described in.

2.3 Shape modeling - statistical atlas generation

We generate the statistical atlases by iteratively averaging subject scans that are mapped onto a common reference space. We first choose one subject to serve as the reference scan. Then, we linearly register all subjects to the selected reference and compute the group average. In the second iteration, we use as reference the average image produced by linear group registration. Next, we map all subjects to the new computed template using nonlinear registration, and compute the average. In the remaining iterations we only apply nonlinear mappings to update the average image. We repeat the above steps until the final iteration $i = K$. This process converged to an atlas template within $K = 5$ iterations on our data.

The above steps generate sequences of transformations $\tau_i^{(n)}$, where $n = 0, 1, 2, \ldots, N$ is the nth image and $i = 0, 1, \ldots, K$ is the ith iteration. Rohlfing et al., 2001 showed that in each iteration i, the transformation $\tau_i^{(n+1)}$ and the preceding transformation $\tau_i^{(n)}$ differ only by a small amount of deformation. Finally, a human operator labeled the trabecular bone, cortical bone, muscle, and SAT using the MIPAV software suite (McAuliffe et al., 2001) by manual selection of control points and spline interpolation. In the first row of Figure 2, we display our computed atlas image and the atlas label map for 4%, 38% and 66% of the tibia length, respectively. In the first row, third, sixth and ninth columns of Figure 2, we display the color-coded atlas labels, for each tibial site under consideration.

2.4 Multi-atlas based tissue segmentation

2.4.1 Label propagation

Label propagation is the process of assigning labels from the warped atlas labels to the reference space. We use the linear (μ) and nonlinear (τ) transformations between S and A that we found in the registration stage, to map labels from the atlas to the subject space. Label propagation is achieved by nearest neighbor interpolation after warping the atlas label to the subject domain *via* the estimated deformation $\tau := \tau \circ \mu$. The segmentation map is produced by $\tau(SA)$.

2.4.2 Label fusion

This is a key stage of MAIS. Here, we combine all the propagated atlases to obtain a final segmentation (Iglesias and Sabuncu 2015). Various methods have been proposed for this stage including best atlas selection, a selective and iterative method for performance level estimation (SIMPLE), joint label fusion, majority voting, weighted majority voting, and simultaneous truth and performance level estimation (STAPLE) algorithm (Warfield et al., 2004; Langerak et al., 2010; Wang et al., 2013; Iglesias and Sabuncu 2015). In this work, we utilized STAPLE for fusing segmentation results by individual atlases. We utilize STAPLE for label fusion because it has performed very well over a range of applications Cardoso et al., 2013; Weston et al., 2019.

STAPLE can be formulated using probabilistic classification terms. Given K segmentations (classifications) of a subject S having N pixels, let $e_k(\mathbf{x})$ be the decision of classifier k at voxel $\mathbf{x}$. If the (unknown) ground truth label for voxel $\mathbf{x}$ is l, we say that $\mathbf{x} \in C_l$. The performance of classifier k is determined by two parameters p (sensitivity) and q (specificity), referring to the fractions of true positives and true negatives among the classified voxels, that maximizes the complete log likelihood function, $(p, q) = \arg\max_{p,q} \ln f\,(D, T|p, q)$, where $T = \cup_{l=1}^{n} C_l$ is the true segmentation, also called missing or hidden data, and $D = [e_k(\mathbf{x})]$ is an $N \times K$ decision matrix.

For each classifier k, and each class C_l, the parameters p and q are modeled independently as the following conditional probabilities: $p_k = \Pr\,(e_k(\mathbf{x}) = l|x \in C_l)$and $q_k = \Pr\,(e_k(\mathbf{x}) \neq l|\mathbf{x} \notin C_l)$. The process of estimation of p and q is achieved by the Expectation-Maximization (EM) algorithm.

The final segmentation $\hat{S}$ at voxel $\mathbf{x}$ is computed by $E(\mathbf{x}) = \arg\max_i P\,(x \in C_i|e_1(\mathbf{x}), \ldots, e_K(\mathbf{x}))$. The probability $p\,(\mathbf{x} \in C_i|e)$ follows from the classifier's decisions and their performance parameters using Bayes' rule (Rohlfing et al., 2005).

3 Data description and performance evaluation measures

3.1 Overview of dataset

We used pQCT data obtained from the InCHIANTI clinical study to evaluate the performance of the methods. InCHIANTI is a longitudinal study of risk factors for mobility disability performed in a representative sample of the middle aged and older populations living in Tuscany, Italy (Ferrucci et al., 2000; Lauretani et al., 2008; Makrogiannis et al., 2018). The validation dataset is randomly sampled from the original InCHIANTI baseline database as in Makrogiannis et al., 2018. It consists of pQCT scans of the lower leg acquired at the 4%, 38% and 66% tibial length of the lower leg. Our randomly sampled dataset contains a total of 77 samples, that is 30 samples at 4%, 27 samples at 38% and 20 samples at 66% tibial length. The pQCT scans of the right lower leg were acquired using a XCT 2000 scanner (Stratec Medizintechnik GmbH, Pforzheim, Germany). The slice thickness of each scan is 2.1 *mm* and the in-plane slice is 0.5 *mm*. At each tibial location, we used three templates for MAIS. One of these three samples was used to initialize statistical atlas generation. We cross-validated the performance of each method on the testing samples that

10.3389/fphys.2022.951368

remained after removing the three templates. We focused on specific tissue(s) of interest at each tibial location: trabecular bone at 4%; cortical and trabecular bone at 38%; subcutaneous adipose tissue (SAT), muscle, cortical bone and trabecular bone at 66% as shown in Table 1.

3.2 Motion artifacts

Unique characteristics of the test data that made it suitable for our analysis and clinically relevant are: 1) the strong representation of older persons who are likely to experience walking difficulties, 2) the different locations of the tibial length that we considered for identification of six different tissue types (Table 1), 3) the incidence of motion artifacts in the dataset making it complicated for segmentation. Motion artifacts occur as a result of the voluntary and involuntary movement of the subject during image acquisition.

Frequent criteria for motion assessment are discontinuities and streaks in the cortical bone, and blurring and shifting of trabecular bone. A clinical specialist used the above visual inspection criteria and a grading system from 1 to 5 (1: no artifacts) to classify the subject motion artifacts as described in Wong 2016. This effect was more evident at the 66% tibial length with more than 50% of the samples having significant motion artifacts corresponding to grades 4 and 5.

3.3 Validation

We performed quantitative analysis of the performance of the atlas-based tissue identification schemes by calculating the Dice Similarity Coefficient (DSC), sensitivity or True Positive Rate (TPR), and Precision (PR) between segmentation results and their ground truths T. We cross-validated the performance metrics by excluding the two subjects we used as templates in MAIS, and the subject we used as template for statistical atlas generation. We made comparisons across single- and multi-atlas based image segmentation methods with respect to registration algorithms. We compared the results obtained by multi-atlas- based techniques with those obtained by the automated tissue identification and quantification (TIDAQ) method (Makrogiannis et al., 2018), to emphasize the performance of MAIS techniques. We chose TIDAQ for our comparisons because it is a model-based tissue segmentation method that has produced good results for images of good quality, but may be challenged by images that have moderate to significant motion artifacts. We performed nonparametric Wilcoxon rank sum tests to examine the effects of motion artifacts on tissue identification using pQCT scans of 66% tibia. The sample size is $n = 20$.

TABLE 2 Summary of the unlabeled pQCT scans of the lower leg from the InCHIANTI Study.

Site (%)	Gender	*N*	Age (yrs)				
			Min	*Max*	*Med*	*Ave*	*Std*
4	Male	272	28	104	74	87.3	14.2
	Female	313	26	104	74	87.4	15.0
	Total	585	26	104	73	87.4	14.6
38	Male	272	28	104	73	69.2	14.5
	Female	311	26	104	74	69.4	14.7
	Total	583	26	104	74	69.3	14.6
66	Male	270	28	104	73	69.2	14.3
	Female	304	26	104	73	69.1	14.8
	Total	580	26	104	73	69.1	14.5

3.4 Tissue quantification

We provide an extension of our analysis to non-labeled pQCT scans of the lower leg from the InCHIANTI dataset to quantify body composition changes caused by aging. Our aim is to show the reliability of the MAIS-technique using SDD-STPL for tissue quantification on an extended clinical dataset and evaluate the agreement of our results with clinical observations. We decided to quantify the baseline InCHIANTI dataset that includes a total of 2,425 pQCT scans. We applied the following procedures to prepare the data for analysis. With the help of TIDAQ software, we sorted the scans according to anatomical sites, selected scans at 4%, 38%, and 66% of tibial length and removed scans with different orientations. We then separated the remaining scans according to their gender (male and females). After data preparation, we had a total of 1748 scans for quantification that may be grouped as follows: 585 scans (males: 272, and females: 313) at 4%; 583 scans (males: 272, females: 311) at 38%; and 580 scans (Males: 270, females: 304) at 66%. A summary of our quantification dataset, including gender and age distribution, is provided in Table 2.To characterize the effect of aging on body composition, we calculated the cross-sectional area (*CSA*) and density of the trabecular bones at 4% tibia, cortical bones at 38% tibia, and muscle and subcutaneous adipose tissue (SAT) at 66% tibia. To obtain these measures, we ran our SDD-STPL technique on all datasets to automatically identify these tissues and computed the total CSA and the average density for each tissue type. We decided to perform analyses of both genders jointly, as well as separate gender-conditional analyses.

We used scatter plots and regression analyses of quantification results to study the changes in body composition (response variable) with respect to age (predictor variable). We utilized the following statistical measures to analyze the relationship between the two variables: the square of the correlation coefficient (R^2), the coefficient of variation of

TABLE 3 4% Trabecular Bone Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Method	DSC	TPR	PR
STAT-FFD	0.947 ± 0.045	0.96 ± 0.021	0.939 ± 0.084
STPL-FFD	0.936 ± 0.056	0.971 ± 0.018	0.911 ± 0.103
STAT-SDD	0.959 ± 0.019	0.939 ± 0.037	0.981 ± 0.015
STPL-SDD	**0.972 ± 0.009**	0.965 ± 0.021	**0.98 ± 0.014**
STAT-SyNO	0.906 ± 0.072	0.922 ± 0.081	0.9 ± 0.103
STPL-SyNO	0.914 ± 0.068	0.936 ± 0.07	0.902 ± 0.104
STAT-SyN	0.947 ± 0.05	0.972 ± 0.017	0.93 ± 0.122
STPL-SyN	0.947 ± 0.054	**0.975 ± 0.02**	0.928 ± 0.128
TIDAQ	0.941 ± 0.022	0.913 ± 0.038	0.972 ± 0.024

the root mean squared difference (*CV-RMSD*) (Eq. 11) between the reference (y) and the predicted ($\hat{y}$) measurements, the slope and intercept of the regression line, and the p − values.

$$CV - RMSD = \frac{1}{\mu}\sqrt{\frac{\sum_{k=1}^{N}(y_k - \hat{y}_k)^2}{N}}; \quad \mu = \frac{\sum_{k=1}^{N} y_k}{N} \tag{11}$$

4 Experiments and results

Here we evaluate the performances of the four MAIS techniques and an automated model-based tissue quantification method (TIDAQ) (Makrogiannis et al., 2018). We then explore the effect of subject movement artifacts on tissue segmentation performance. Finally, we expand our analysis to quantify the complete InCHIANTI baseline dataset.

Our aim is to support or reject the hypotheses that 1) MAIS techniques, in general, improve the segmentation performance of SAIS, 2) STAPLE on SDD mappings produces better segmentation quality than the other methods, 3) MAIS is more resilient to subject movement than model-based segmentation.

We analyzed the performance of the deformable methods by validating the SAIS results of statistical atlases against reference standards, T. Reference standard is a tissue label map that was generated manually by a clinical specialist. We evaluated single atlas segmentation performances of STAT-FFD, STAT-SDD, STAT-SyN, and STAT-SyNO, where 'STAT' represents the statistical atlas. The MAIS techniques we developed and evaluated in this framework are STPL-FFD, STPL-SDD, STPL-SyN, and STPL-SyNO. 'SyNO' denotes an 'Symmetric Normalization with our own linear registration' and 'STPL' denotes 'STAPLE.'

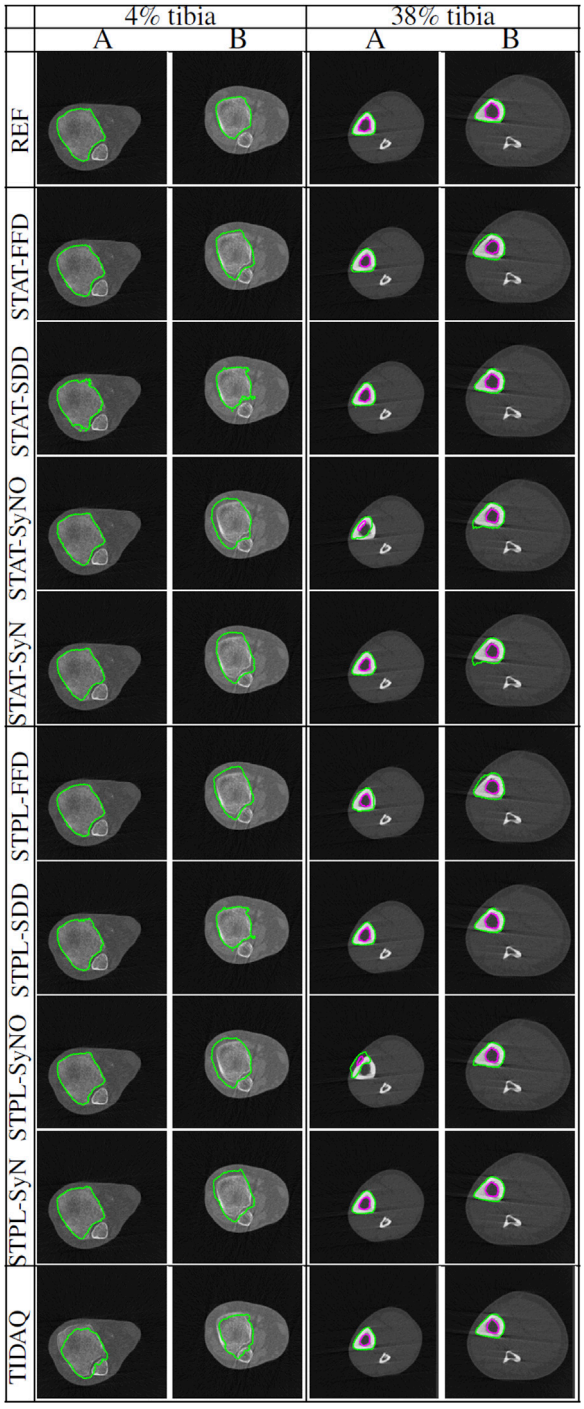


FIGURE 4
Comparisons of tissue segmentation by MAIS and model-based methods at 4% tibia and 38% tibia. Two subjects (subject **(A)** and subject **(B)**) are selected at each tibial site to demonstrate the performance of the compared methods. At 4% the trabecular bone is delineated by the green contour, and at 38% the cortical bone and the trabecular bone are delineated by the green and magenta contours, respectively.

10.3389/fphys.2022.951368

TABLE 4 38% Cortical Bone Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Method	DSC	TPR	PR
STAT-FFD	0.809 ± 0.126	0.933 ± 0.064	0.721 ± 0.125
STPL-FFD	0.781 ± 0.136	0.889 ± 0.135	0.706 ± 0.153
STAT-SDD	0.932 ± 0.022	0.965 ± 0.025	0.902 ± 0.032
STPL-SDD	**0.947 ± 0.021**	0.949 ± 0.036	**0.947 ± 0.023**
STAT-SyNO	0.78 ± 0.174	0.891 ± 0.173	0.704 ± 0.184
STPL-SyNO	0.76 ± 0.215	0.841 ± 0.233	0.704 ± 0.216
STAT-SyN	0.871 ± 0.07	**0.983 ± 0.026**	0.787 ± 0.104
STPL-SyN	0.873 ± 0.112	0.96 ± 0.079	0.807 ± 0.141
TIDAQ	0.897 ± 0.15	0.954 ± 0.156	0.847 ± 0.147

TABLE 5 38% Trabecular Bone Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Method	DSC	TPR	PR
STAT-FFD	0.823 ± 0.084	0.774 ± 0.145	0.917 ± 0.121
STPL-FFD	0.797 ± 0.131	0.759 ± 0.177	0.891 ± 0.155
STAT-SDD	0.92 ± 0.027	0.868 ± 0.052	**0.981 ± 0.019**
STPL-SDD	**0.939 ± 0.019**	**0.91 ± 0.041**	0.972 ± 0.026
STAT-SyNO	0.703 ± 0.294	0.627 ± 0.282	0.831 ± 0.309
STPL-SyNO	0.687 ± 0.343	0.624 ± 0.338	0.807 ± 0.363
STAT-SyN	0.849 ± 0.079	0.764 ± 0.081	0.97 ± 0.109
STPL-SyN	0.877 ± 0.08	0.811 ± 0.11	0.972 ± 0.078
TIDAQ	0.825 ± 0.167	0.726 ± 0.156	0.96 ± 0.187

4.1 4% tibia segmentation

Table 3 displays the summarized performance measures obtained by the single-atlas image segmentation using the statistical atlas and multi-atlas image segmentation methods, and TIDAQ for identification of the trabecular bone (TB) over the test-set. In Figure 4, column 1 we show examples of multi-atlas segmentations of the trabecular bone at 4% tibia by the compared methods, delineated by the green contours.

A comparison of single-atlas and multi-atlas segmentation results in Table 3 shows that the use of multiple atlases improves segmentation quality. With respect to multi-atlas image segmentation techniques, we examined the performance quality of the methods and the effect of label fusion on the improvement of the results. The range and mean ± standard deviation of DSC values produced by STPL-SDD are [0.944, 0.992] and 0.972 ± 0.009, by STPL-FFD are [0.739, 0.991] and 0.936 ± 0.056, by STPL-SyN are [0.743, 0.990] and 0.947 ± 0.054, and by STPL-SyNO are [0.731, 0.991] are 0.914 ± 0.068, respectively, as can be seen in Table 3. In summary, we observe that STPL-SDD outperformed the other multi-atlas techniques.

The range of DSC values produced by TIDAQ for trabecular bone identification is [0.897, 0.967] with mean ± standard deviation of 0.941 ± 0.022. The results in Table 3 indicate that TIDAQ outperformed STPL-FFD, and STPL-SyNO registration techniques. On the other hand, TIDAQ was less accurate than STPL-SDD, and STPL-SyN.

4.2 38% tibia segmentation

Tables 4 and 5 contain the performance measures produced by these experiments. Figure 4 displays examples of tissue delineations, on two of our test subjects.

Overall, quantitative results of the methods reported in Tables 4 and 5, show that SDD produced better segmentation quality than the other deformable models, followed by SyN. Low DSC values are produced by SyN, FFD, and SyNO models (in decreasing order) for both cortical and trabecular bone in some subjects. SyNO missed the trabecular bone in few subjects producing zero DSC. We also noticed that all deformable methods produced DSC means that are greater than 75% for cortical and trabecular bone except SyNO in the trabecular bone. We infer that the range of DSC values for SDD is more compact than the other deformable models. The minimum value of DSC produced by SDD is greater than 85% for cortical bone, and about 82% for trabecular bone. The DSC values produced by SyN are fairly compact in the identification of bone compartments, and their mean values are greater than 85% with standard deviations about 10%.

In Tables 4 and 5, we observe that FFD results for both single- and multi-atlas image segmentation techniques show a wider DSC dispersion than SDD, the former producing values lower than 60% for cortical bone and about 45% for trabecular bone.

SyNO results show wider DSC spread than all of the other deformable models. SyNO produced the least mean DSC of about 78% for identification of cortical bone and about 68% for trabecular bone. This error is usually caused by linear registration failures propagated to the symmetric normalization stage.

TIDAQ performed very well in the identification of trabecular bone, and much better in the identification of cortical bone. Overall, the identification accuracy is promising with DSC mean ± standard deviation of 89.7 ± 2.2% for cortical bone and 82.5 ± 16.7% for trabecular bone. Despite the good performance of TIDAQ, STPL-SDD outperformed it in the identification of cortical bone, while STPL-SDD and STPL-SyN outperformed it in the identification of trabecular bone. This shows that multi-atlas image segmentation techniques have the potential to produce higher tissue identification accuracy than TIDAQ.

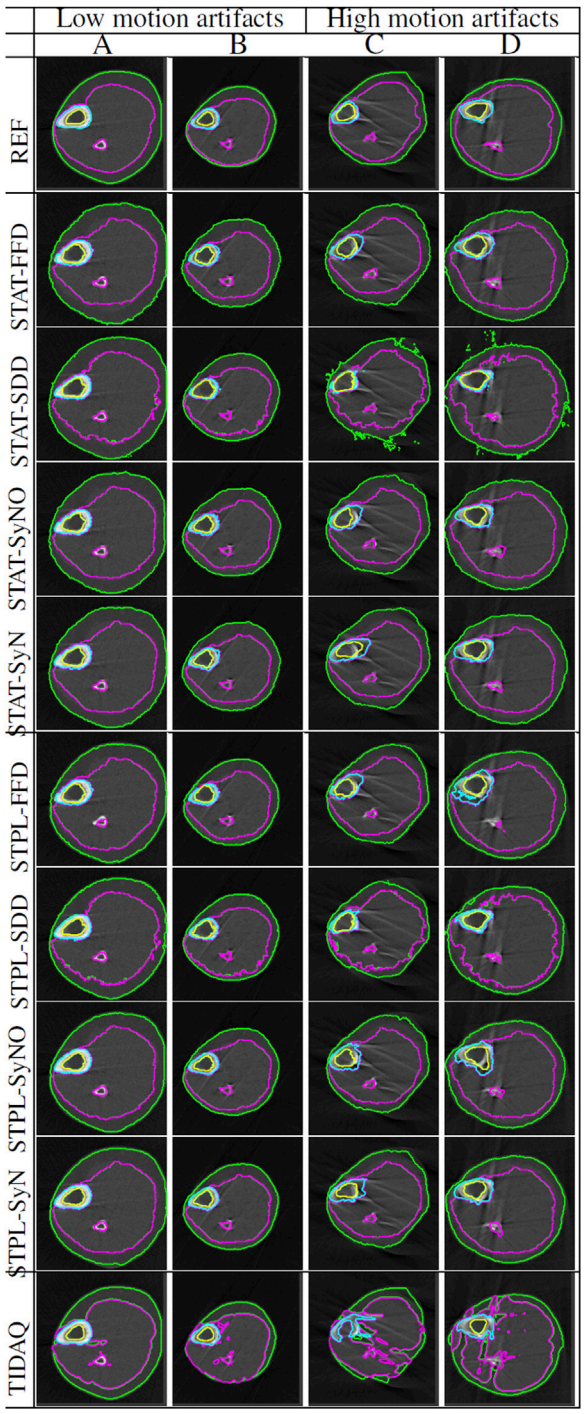


FIGURE 5
Segmentation comparisons of scans at 66% tibial length with low and high levels of artifacts caused by subject motion. Scans **(A,B)** have low motion artifacts, while scans **(C,D)** have high motion artifacts. The delineation of the subcutaneous adipose tissue (SAT) is represented by the green contour, muscle by magenta color, cortical bone by cyan color, and trabecular bone by yellow color.

TABLE 6 66% SAT Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.682 ± 0.158	**0.953 ± 0.051**	0.562 ± 0.2
STPL-FFD	0.73 ± 0.16	0.912 ± 0.085	0.649 ± 0.227
STAT-SDD	0.708 ± 0.151	0.939 ± 0.042	0.594 ± 0.189
STPL-SDD	**0.771 ± 0.144**	0.917 ± 0.057	0.692 ± 0.202
STAT-SyNO	0.684 ± 0.152	0.938 ± 0.065	0.572 ± 0.2
STPL-SyNO	0.729 ± 0.161	0.892 ± 0.093	0.659 ± 0.229
STAT-SyN	0.677 ± 0.147	0.935 ± 0.079	0.565 ± 0.197
STPL-SyN	0.734 ± 0.157	0.889 ± 0.078	0.664 ± 0.225
TIDAQ	0.746 ± 0.228	0.845 ± 0.14	**0.702 ± 0.267**

TABLE 7 66% Muscle Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.893 ± 0.032	0.837 ± 0.067	0.962 ± 0.037
STPL-FFD	0.902 ± 0.034	0.857 ± 0.074	0.959 ± 0.037
STAT-SDD	0.922 ± 0.029	0.891 ± 0.057	0.958 ± 0.032
STPL-SDD	**0.938 ± 0.028**	**0.914 ± 0.057**	**0.966 ± 0.029**
STAT-SyNO	0.894 ± 0.038	0.847 ± 0.073	0.953 ± 0.054
STPL-SyNO	0.911 ± 0.038	0.874 ± 0.078	0.959 ± 0.048
STAT-SyN	0.893 ± 0.035	0.844 ± 0.077	0.957 ± 0.051
STPL-SyN	0.913 ± 0.04	0.875 ± 0.084	0.962 ± 0.045
TIDAQ	0.904 ± 0.085	0.861 ± 0.131	0.963 ± 0.025

4.3 66% tibia segmentation

Figure 5 displays segmentation results produced by the compared approaches on two scans with low motion degradation and two scans with high motion degradation. The mean ± standard deviation values of DSC, TPR and PR of each tissue over all testing samples are given in Tables 6–9. Overall, SDD exhibited better performance across the three metrics than the other deformable models for the identification of all tissues, followed by SyN and TIDAQ. In addition, SDD produced DSC, true positive rate, and precision values of lower dispersion (expressed by smaller standard deviation) than the other methods. STPL-SDD yielded the top DSC performance for SAT, muscle, and trabecular bone segmentation. The DSC minimum values for this method were about 62.7% for SAT, 89% for muscle, 75% for cortical bone, and 86.2% for trabecular bone. All MAIS techniques produced mean DSC greater than 89%, mean TPR greater than 84%, and mean PR greater than 95% in muscle segmentation. Conversely, all tested methods yielded

TABLE 8 66% Cortical Bone Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.665 ± 0.152	0.668 ± 0.121	0.677 ± 0.196
STPL-FFD	0.668 ± 0.132	0.696 ± 0.11	0.666 ± 0.19
STAT-SDD	0.681 ± 0.094	0.55 ± 0.112	**0.911 ± 0.059**
STPL-SDD	0.829 ± 0.083	0.776 ± 0.118	0.898 ± 0.064
STAT-SyNO	0.721 ± 0.156	0.708 ± 0.145	0.754 ± 0.201
STPL-SyNO	0.765 ± 0.194	0.792 ± 0.181	0.761 ± 0.225
STAT-SyN	0.718 ± 0.147	0.719 ± 0.084	0.737 ± 0.21
STPL-SyN	0.788 ± 0.152	0.827 ± 0.112	0.769 ± 0.194
TIDAQ	**0.851 ± 0.115**	**0.987 ± 0.032**	0.76 ± 0.157

TABLE 9 66% Trabecular Bone Segmentation Performance (mean ± standard deviation). DSC: Dice Similarity Coefficient, TPR: True Positive Rate, PR: Precision.

Methods	DSC	TPR	PR
STAT-FFD	0.825 ± 0.087	0.902 ± 0.124	0.783 ± 0.124
STPL-FFD	0.827 ± 0.085	0.911 ± 0.123	0.783 ± 0.138
STAT-SDD	0.86 ± 0.049	**0.983 ± 0.023**	0.769 ± 0.08
STPL-SDD	**0.913 ± 0.045**	0.968 ± 0.037	0.869 ± 0.088
STAT-SyNO	0.843 ± 0.179	0.91 ± 0.203	0.803 ± 0.142
STPL-SyNO	0.868 ± 0.167	0.913 ± 0.191	0.838 ± 0.14
STAT-SyN	0.856 ± 0.116	0.946 ± 0.073	0.791 ± 0.149
STPL-SyN	0.895 ± 0.105	0.939 ± 0.077	0.862 ± 0.133
TIDAQ	0.871 ± 0.221	0.816 ± 0.216	**0.994 ± 0.026**

mean DSC less than 80%, and a top PR of 70.2% in SAT segmentation.

In SAT delineation, FFD and SyNO models compete with each other in the identification of SAT, but SyNO outperforms free-form deformation in the identification of other tissues by at least 10% accuracy. Except for cortical bone, we observe that MAIS outperformed TIDAQ.

4.4 Effects of subject motion on tissue identification at 66% tibia

In this experiment, we studied the effect of artifacts on segmentation performance. The motion was assessed using 5-level visual grading as described in Wong (2016). We separated the samples into low level of motion determined by grades 1–3, and high level of motion with grades 4 and 5, and compared the performance of all methods. As stated above, 12 out of 20 pQCT scans at 66% tibia contain high to severe motion artifacts (grades 4 and 5).

The next step is to explore the differences in segmentation performance between the two groups. Table 10 summarizes the segmentation performance measured by DSC, true positive rate (TPR) and precision (PR) for each tissue type and each method. This table also contains the relative differences of the performance measures. To estimate the statistical significance of the differences in segmentation performance, we applied nonparametric Wilcoxon rank sum tests between the two groups and we report the *p*-values.

Considering SAT, we observe consistent decrease of average DSC and PR with increasing motion artifacts for all methods. We observe that TIDAQ shows the highest decrease in all performance measures. The Wilcoxon tests indicate statistically significant performance differences for TIDAQ in DSC and precision, and for STPL-SDD in DSC and true positive rate.

4.5 Tissue composition assessment

We applied our STPL-SDD method to the baseline InCHIANTI dataset that we described in Section 3 and summarized in Table 2. We then analyzed the quantification results of cross-sectional areas and average densities for each tissue to explore changes in its composition as a function of age. The scatter plots and regression results in Figures 6, 7 lead to the following observations. Trabecular bone density decreases with age at similar rates for males and females. The CSA of cortical bone decreases with age at similar rates for males and females. Cortical bone density decreases with age for males and females, and the rate of decrease is higher for females. Muscle CSA decreases with age for males and females, and the rate of decrease is higher for males. Muscle density decreases with age at similar rates for males and females. Our analysis does not reveal significant changes with age for SAT CSA, SAT density, and trabecular bone CSA.

In addition, the statistical results reported in Tables 11–13 show that there is increased correlation of cortical bone density with age, and trabecular bone density with age, especially for females, relative to the other tissues. In addition, there is noticeable correlation between muscle CSA and age for males. The CV-RMSD values show decreased variation mostly for cortical density and muscle density for each gender. We also observe that the gender-conditional analyses produce lower variability than joint analyses of males and females as we expected.

4.6 Method implementation and execution time

We developed the programs of the proposed methodologies in C++, *Python* 3.7, and used the ITK library. We implemented

TABLE 10 Effect of motion artifacts on segmentation performance.

Dice similarity coefficient (DSC)

TISSUE	66%-SAT				66%-MUSCLE				66%-CB				66%-TB			
MOTION	Low	High	Diff%	*p*	Low	High	Diff%	*p*	Low	High	Diff%	*p*	Low	High	Diff%	*p*
STPL-FFD	0.799	0.703	−11.9%	0.208	0.919	0.896	−2.5%	0.143	0.691	0.660	−4.5%	0.849	0.785	0.843	7.4%	0.173
STPL-SDD	0.866	0.734	−15.2%	**0.046**	0.964	0.928	−3.8%	**0.004**	0.876	0.810	−7.5%	0.059	0.908	0.914	0.7%	0.849
STPL-SyN	0.794	0.711	−10.5%	0.246	0.926	0.908	−2.0%	0.503	0.792	0.786	−0.8%	0.173	0.845	0.914	8.2%	0.453
STPL-SyNO	0.814	0.697	−14.4%	0.173	0.936	0.901	−3.7%	0.095	0.813	0.747	−8.1%	0.143	0.900	0.856	−4.9%	0.775
TIDAQ	0.922	0.678	−26.4%	**0.014**	0.978	0.876	−10.5%	$4.7{\cdot}10^{-4}$	0.956	0.811	−15.2%	$2.3{\cdot}10^{-4}$	0.960	0.837	−12.8%	**0.035**

TRUE POSITIVE RATE (TPR)

TISSUE	66%-SAT				66%-MUSCLE				66%-CB				66%-TB			
MOTION	Low	High	Diff%	*p*	Low	High	Diff%	*p*	Low	High	Diff%	*p*	Low	High	Diff%	*p*
STPL-FFD	0.914	0.912	−0.3%	0.703	0.882	0.847	−4.0%	0.443	0.695	0.697	0.2%	0.633	0.952	0.895	−6.0%	0.633
STPL-SDD	0.958	0.901	−5.9%	**0.007**	0.950	0.901	−5.1%	0.117	0.819	0.760	−7.2%	0.336	0.966	0.969	0.3%	0.387
STPL-SyN	0.886	0.890	0.5%	0.775	0.894	0.868	−2.9%	0.775	0.822	0.829	0.8%	0.849	0.925	0.944	2.1%	1.000
STPL-SyNO	0.903	0.888	−1.6%	0.703	0.911	0.860	−5.7%	0.246	0.843	0.772	−8.4%	0.387	0.980	0.887	−9.4%	0.143
TIDAQ	0.911	0.820	−10.0%	0.336	0.978	0.817	−16.4%	$2.3{\cdot}10^{-4}$	1.000	0.982	−1.8%	0.185	0.925	0.774	−16.3%	**0.035**

PRECISION (PR)

TISSUE	66%-SAT				66%-MUSCLE				66%-CB				66%-TB			
MOTION	Low	High	Diff%	*p*	Low	High	Diff%	*p*	Low	High	Diff%	*p*	Low	High	Diff%	*p*
STPL-FFD	0.732	0.617	−15.7%	0.503	0.964	0.957	−0.6%	0.924	0.719	0.645	−10.3%	0.387	0.672	0.826	23.0%	**0.026**
STPL-SDD	0.797	0.651	−18.3%	0.289	0.980	0.960	−2.0%	0.173	0.958	0.875	−8.7%	**0.002**	0.868	0.870	0.3%	0.775
STPL-SyN	0.747	0.631	−15.5%	0.443	0.966	0.961	−0.6%	0.775	0.794	0.759	−4.4%	0.117	0.791	0.890	12.6%	0.443
STPL-SyNO	0.766	0.617	−19.4%	0.246	0.967	0.956	−1.1%	1.000	0.814	0.740	−9.1%	0.117	0.838	0.839	0.2%	0.566
TIDAQ	0.933	0.613	−34.3%	**0.010**	0.979	0.957	−2.2%	0.117	0.915	0.700	−23.5%	$2.3{\cdot}10^{-4}$	1.000	0.992	−0.8%	1.000

the symmetric normalizations, -SyN and -SyNO (originally -SyNOnly) of ANTs on the ANTsPY python library (antspyx version 0.2.5) with default parameters to generate the segmentation results corresponding to these nonlinear registration algorithms. The TIDAQ backend is implemented in C++ and uses ITK, while the user interface is a Java plugin. We executed our experiments on a system with Linux CentOS 7, 2 x Intel(R) Xeon(R) CPU E5-2,690 v4 2.60 GHz, and 128 GB RAM. We computed the execution time of SAIS and MAIS, with respect to the deformable registration algorithms, for all subjects at the different tibia location. We calculated the mean ± standard deviation of the execution time over all subjects in the segmentation set and report the values in Table 14. We observe that the execution time of STAT-SDD is about 27 s for 4% tibia, and about 32 s for 38% and 66% tibia, while the execution time of STPL-SDD is about 102 s for 4%, 118 s for 38%, and 148 s for 66%.

5 Discussion

5.1 Method comparisons, tissue separability and technical characteristics

Tissue delineation is an important, but challenging task in medical image analysis. The accuracy of tissue delineation

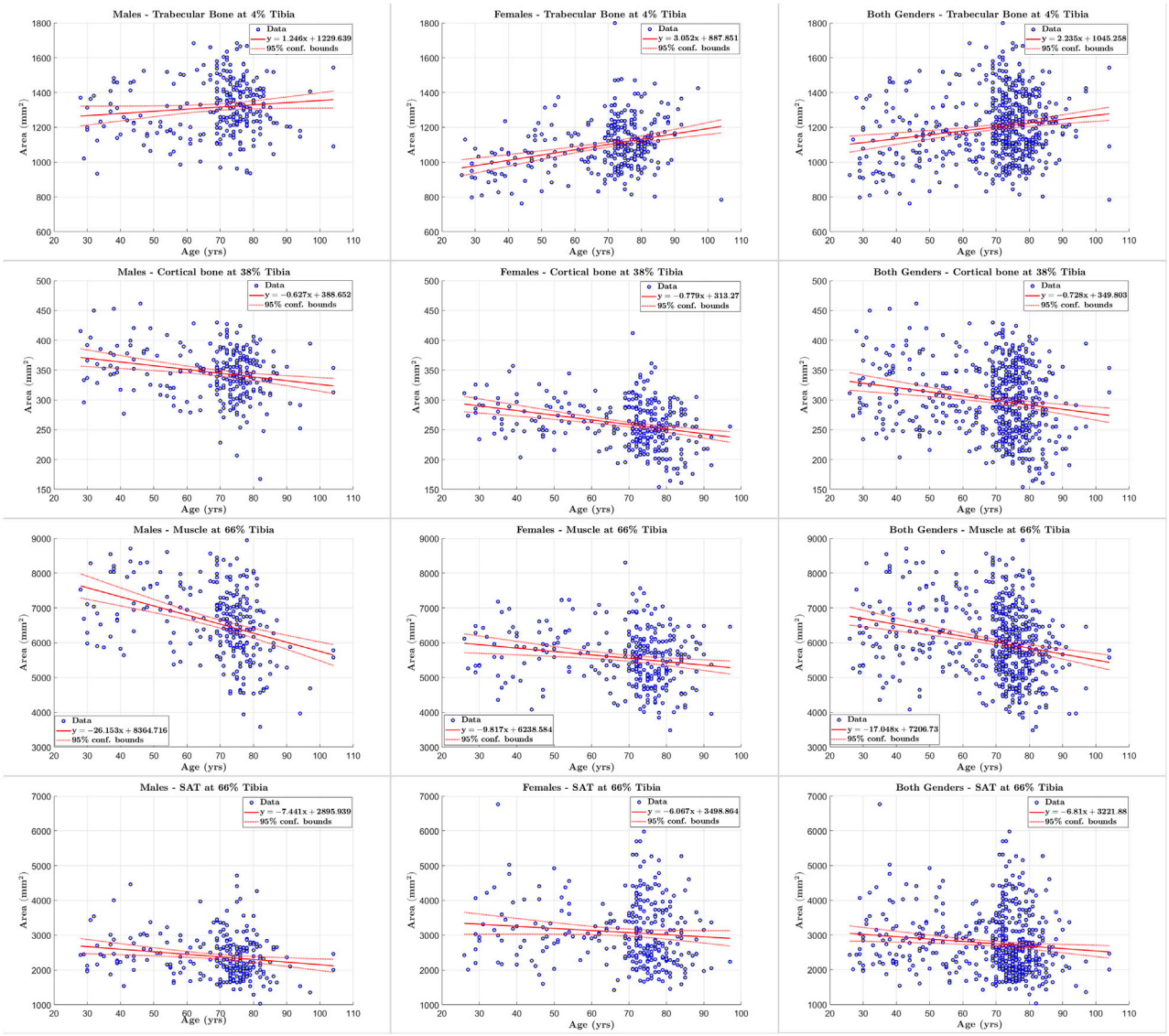


FIGURE 6
Scatter plots of tissue of interest's cross sectional area *versus* age for males, females, and both genders at 4%, 38% and 66% tibia.

depends on many factors, including the intrinsic characteristics of the technique, image modality, artifacts or noise, and the number of tissues to be identified in each scan. Peripheral quantitative computed tomography imaging quality is significantly affected by artifacts caused by subject movement. Overall, the proposed multi-atlas image segmentation techniques address the aforementioned factors. Furthermore, the multi-atlas symmetric diffeomorphic demons technique proved to be more robust to reduced image quality than the other methods, followed by symmetric image normalization.

Visual inspection of tissue densities in the image at multiple tibial sites that are displayed in Figure 8, second row, shows that the distributions of different tissues significantly overlap with one another. At 4% tibia, we observe a clear overlap between the distribution of all leg tissues and the trabecular bone. An optimized thresholding technique based on tissue densities, for example, may not separate the distributions accurately, because of high false positives and false negatives. At 38% tibia, although it appears that there is a valley between the cortical and trabecular bones, yet there is still significant overlap between the distributions of the trabecular bone tissues and all leg tissues. At 66% tibia, we note the extensive overlap among the distributions of SAT, muscle, and trabecular bone. These distributions illustrate the difficulties that would be

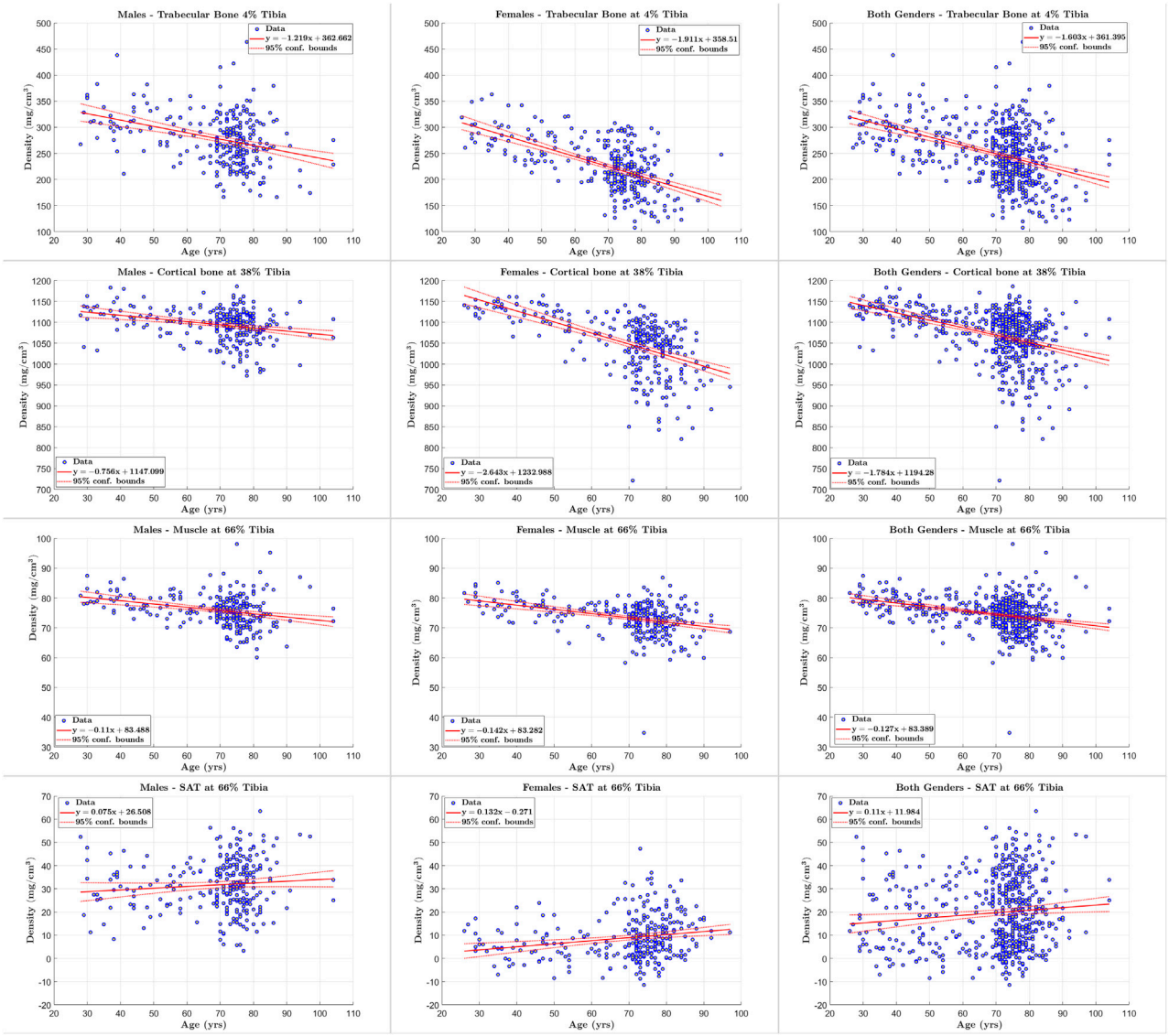


FIGURE 7
Scatter plots of tissue of interest's density *versus* age for males, females, and both genders at 4%, 38% and 66% tibia.

encountered by segmentation techniques that rely on density alone. Yet, the proposed MAIS technique STPL-SDD, yields high segmentation accuracy (DSC >90%), in almost all tissue delineations at different sites, except for 66%-SAT and 66%-cortical bone.

Specific properties of these methods that improved the tissue delineation accuracy are their 1) symmetric nature, and 2) diffeomorphism. Concerning symmetry, the method in question provides equal treatment to both fixed and moving images. In addition, the interactive force between the two images can produce accurate registration of one part of the image to the other, and *vice versa* (Rogelj and Kovačič, 2006). On the other hand, diffeomorphism affords the algorithms the ability to handle both large and small deformations (Sotiras et al., 2013). It is important to note that large deformations are a result of large strains or rotations, which are caused by subject movement. Thus, symmetric diffeomorphic demons and symmetric normalization are robust to local image artifacts or large image deformations that are difficult to register.

The segmentation results shown in Figure 4, at 4% and 38% tibial length, correspond to subjects corresponding to variable levels of delineation challenges, caused by either the condition of the subject (in the 4% examples) or subject motion (in the 38% examples). At both anatomical sites, subject 'B' presents

TABLE 11 Statistical measures of the relationships between tissue properties and age for males and females.

Measurement	Site	R^2	CV-RMSD	Slope	Intercept	p-value
Trabecular CSA	4	0.031	3.646	2.235	1,045.3	$< 10^{-4}$
Trabecular density	4	0.181	4.799	−1.603	361.4	$< 10^{-4}$
Cortical CSA	38	0.033	4.661	−0.728	349.8	$< 10^{-4}$
Cortical density	38	0.178	1.260	−1.784	1,194.3	$< 10^{-4}$
SAT CSA	66	0.013	7.361	−6.810	3,221.9	0.0054
SAT density	66	0.011	18.583	0.110	12	0.0123
Muscle CSA	66	0.058	3.967	−17.048	7,206.7	$< 10^{-4}$
Muscle density	66	0.113	1.663	−0.127	83.4	$< 10^{-4}$

TABLE 12 Statistical measures of the relationships between tissue properties and age for females.

Measurement	Site	R^2	CV-RMSD	Slope	Intercept	p-value
Trabecular CSA	4	0.102	2.159	3.052	887.9	$< 10^{-4}$
Trabecular density	4	0.348	3.025	−1.911	358.5	$< 10^{-4}$
Cortical CSA	38	0.086	2.528	−0.779	313.3	$< 10^{-4}$
Cortical density	38	0.307	0.977	−2.643	1,233	$< 10^{-4}$
SAT CSA	66	0.010	5.122	−6.067	3,498.9	0.0870
SAT density	66	0.044	17.826	0.132	−0.3	0.0002
Muscle CSA	66	0.034	2.428	−9.817	6,238.6	0.0013
Muscle density	66	0.150	1.185	−0.142	83.3	$< 10^{-4}$

TABLE 13 Statistical measures of the relationships between tissue properties and age for males.

Measurement	Site	R^2	CV-RMSD	Slope	Intercept	p-value
Trabecular CSA	4	0.013	1.950	1.246	1,229.6	0.0611
Trabecular density	4	0.127	2.704	−1.219	362.7	$< 10^{-4}$
Cortical CSA	38	0.048	1.937	−0.627	388.7	0.0003
Cortical density	38	0.081	0.554	−0.756	1,147.1	$< 10^{-4}$
SAT CSA	66	0.032	4.019	−7.441	2,895.9	0.0032
SAT density	66	0.009	5.721	0.075	26.5	0.1141
Muscle CSA	66	0.135	2.362	−26.153	8,364.7	$< 10^{-4}$
Muscle density	66	0.087	1.098	−0.110	83.5	$< 10^{-4}$

TABLE 14 Mean ± standard deviation (in seconds) of the execution times of SAIS and MAIS with respect to nonlinear registration models.

Tibia site	Method	FFD	SDD	SyN	SyNO
4%	SAIS	56.47 ± 3.34	27.01 ± 0.81	5.55 ± 0.30	37.25 ± 2.91
	MAIS	191.45 ± 9.54	101.82 ± 1.72	41.27 ± 0.99	136.48 ± 5.61
38%	SAIS	55.49 ± 7.00	31.15 ± 1.46	5.67 ± 0.33	37.16 ± 1.90
	MAIS	189.31 ± 18.62	117.18 ± 3.97	41.88 ± 1.34	140.61 ± 15.20
66%	SAIS	55.47 ± 3.96	31.48 ± 1.29	4.42 ± 4.42	36.38 ± 1.30
	MAIS	209.66 ± 31.41	147.18 ± 36.05	52.51 ± 13.25	151.13 ± 25.72

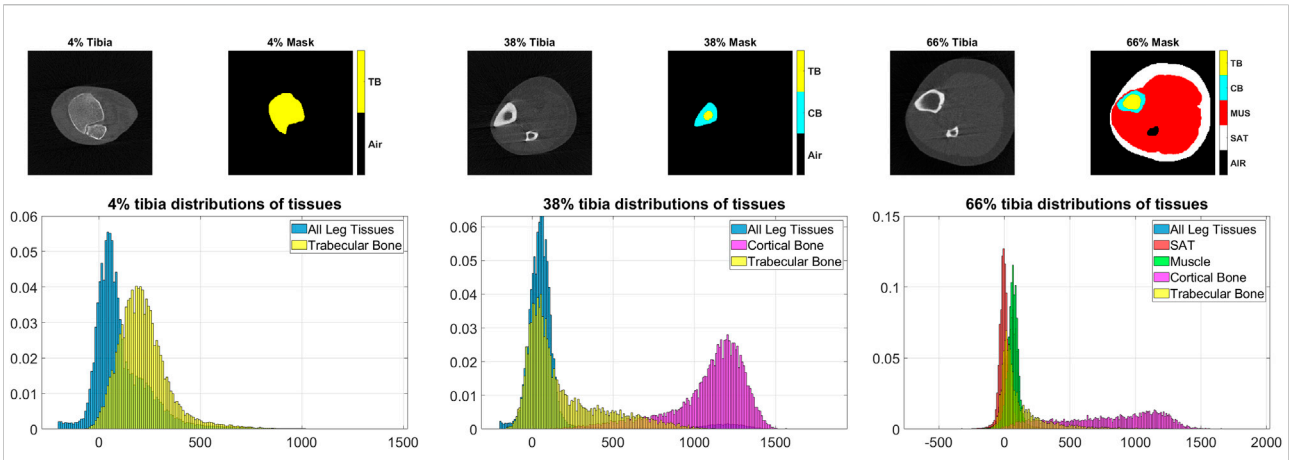


FIGURE 8
Examples of the pQCT scans, masks and tissue distributions. Top row: image-mask pairs of pQCT scans at 4%, 38%, and 66% tibial length respectively. Bottom: distributions of tissues corresponding to the above tibia sites. On the masks, air, trabecular bone, cortical bone, muscle, and subcutaneous adipose tissue (SAT) are identified by black, yellow, cyan, red, and white colors, respectively.

more segmentation challenges than subject 'A'. Visual inspection of subject 'B' at 4% tibia in Figure 4 shows that STPL-FFD and STPL-SyNO could not delineate the trabecular bone accurately when compared to STPL-SDD and STPL-SyN. Similarly, under 38%-tibia scans in Figure 4, we observe that due to the higher presence of streaks in subject 'B' than in subject 'A', STPL-FFD and STPL-SyNO produced lower tissue delineation accuracy than STPL-SDD and STPL-SyN. This observation provides insight into the improved delineation accuracy produced by symmetric diffeomorphic demons and symmetric normalization.

5.2 Effects of subject motion on tissue identification at 66% tibia

The effect of motion artifacts is less pronounced on muscle identification in terms of the relative differences. We observe statistically significant differences in DSC for STPL-SDD and TIDAQ. TIDAQ also showed statistically significant decrease of muscle true positive rate.

In the cortical bone, most performance changes are not statistically significant. TIDAQ produced p-values smaller than 0.05 for DSC and precision, and STPL-SDD for precision only. In the trabecular bone, STPL-FFD produced p-values smaller than 0.05 for precision, and TIDAQ for DSC and true positive rate. DSC and precision values of the trabecular bone produced by STPL-FFD and STPL-SyN increase from low to high motion group, because of segmentation errors in subjects of the low motion group.

We observe that TIDAQ is significantly affected by motion artifacts, as average DSC clearly decreases from the low to high subject motion group. MAIS techniques and especially STPL-SyN are more resilient to subject motion. STPL-SDD still produces the highest DSC and precision values in the high motion group overall.

5.3 Tissue composition assessment

The results in Figures 6, 7 and Tables 11–13 are consistent with findings of clinical studies of aging that used semi-manual quantification workflows (Makrogiannis et al., 2018; Ferrucci et al., 2000). These results indicate that our automated methodology can help to increase the throughput of sophisticated cross-sectional and longitudinal analyses of tissue properties. We also note that the proposed methodology enables the analysis of both hard and soft tissues in pQCT. This is a desirable and innovative feature as pQCT has been mostly restricted to quantification of bone in the past (Gabel et al., 2018; Wong and Manske, 2020). Our methodology opens the door for efficient exploration of muscle properties in the lower leg using pQCT. On the other hand, a greater number of reference segmentation masks may be needed to improve the statistical power of performance evaluations.

6 Conclusion

We introduced multi-atlas segmentation methods for soft and hard tissue segmentation in the lower leg using pQCT data. Our results indicate that the MAIS technique, STPL-SDD, produced more accurate tissue delineation as measured by DSC than all compared methods. STPL-SyN is largely resilient

to subject motion artifacts and noise. The results of our experiments indicate that our methodology can analyze data with degradations caused by subject motion that conventional methods cannot analyze. Future directions of this work include extending this framework to 3D imaging data, and using the segmentation and quantification results for disease prognosis and diagnosis.

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 studies involving human participants were reviewed and approved by the National Institute on Research and Care of the Elderly (INRCA) Institutional Review Board (IRB00001456). The patients/participants provided their written informed consent to participate in this study.

Author contributions

(I) Conception and design: SM, LF; (II) Administrative support: SM, LF; (III) Provision of study materials or patients: SB, LF; (IV) Collection and assembly of data: SM, AO, AD, SB, LF; (V) Data analysis and interpretation: AO, SM, AD; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Funding

This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) (award #SC3GM113754) and by the Army Research Office under grant #W911NF2010095. We acknowledge the support by Delaware CTR-ACCEL (NIH #U54GM104941) and the State of Delaware.

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

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi:10.1016/j.media.2007.06.004

Blew, R. M., Lee, V. R., Farr, J. N., Schiferl, D. J., and Going, S. B. (2014). Standardizing evaluation of pqct image quality in the presence of subject movement: Qualitative versus quantitative assessment. *Calcif. Tissue Int.* 94, 202–211. doi:10.1007/s00223-013-9803-x

Burns, J. E., Yao, J., and Summers, R. M. (2020). Artificial intelligence in musculoskeletal imaging: A paradigm shift. *J. Bone Min. Res.* 35, 28–35. doi:10.1002/jbmr.3849

Cardoso, M. J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., et al. (2013). Steps: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation. *Med. Image Anal.*, 17, 671–684. doi:10.1016/j.media.2013.02.006

Chan, A. C., Adachi, J. D., Papaioannou, A., and Wong, A. K. O. (2018). Investigating the effects of motion streaks on pqct-derived leg muscle density and its association with fractures. *J. Clin. Densitom.* 21, 130–139. doi:10.1016/j.jocd.2016.12.001

Chen, Y., Hu, Y., Yu, Y. E., Zhang, X., Watts, T., Zhou, B., et al. (2018). Subchondral trabecular rod loss and plate thickening in the development of osteoarthritis. *J. Bone Min. Res.* 33, 316–327. doi:10.1002/jbmr.3313

Chow, S. K. H., van Mourik, M., Hung, V. W. Y., Zhang, N., Li, M. M. C., Wong, R. M. Y., et al. (2022). Hr-pqct for the evaluation of muscle quality and intramuscular fat infiltration in ageing skeletal muscle. *J. Pers. Med.* 12, 1016. doi:10.3390/jpm12061016

Cordova, J. S., Schreibmann, E., Hadjipanayis, C. G., Guo, Y., Shu, H. K. G., Shim, H., et al. (2014). Quantitative tumor segmentation for evaluation of extent of glioblastoma resection to facilitate multisite clinical trials. *Transl. Oncol.* 7, 40–47. doi:10.1593/tlo.13835

Ferrucci, L., Bandinelli, S., Benvenuti, E., Di Iorio, A., Macchi, C., Harris, T. B., et al. (2000). Subsystems contributing to the decline in ability to walk: Bridging the gap between epidemiology and geriatric practice in the inchianti study. *J. Am. Geriatr. Soc.* 48, 1618–1625. doi:10.1111/j.1532-5415.2000.tb03873.x

Gabel, L., Macdonald, H. M., Nettlefold, L. A., and McKay, H. A. (2018). Sex-ethnic-and age-specific centile curves for pqct-and hr-pqct-derived measures of bone structure and strength in adolescents and young adults. *J. Bone Min. Res.* 33, 987–1000. doi:10.1002/jbmr.3399

Iglesias, J. E., and Sabuncu, M. R. (2015). Multi-atlas segmentation of biomedical images: A survey. *Med. Image Anal.* 24, 205–219. doi:10.1016/j.media.2015.06.012

Langerak, T. R., van der Heide, U. A., Kotte, A. N., Viergever, M. A., Van Vulpen, M., Pluim, J. P., et al. (2010). Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans. Med. Imaging* 29, 2000–2008. doi:10.1109/TMI.2010.2057442

Lauretani, F., Bandinelli, S., Griswold, M. E., Maggio, M., Semba, R., Guralnik, J. M., et al. (2008). Longitudinal changes in bmd and bone geometry in a population-based study. *J. Bone Min. Res.* 23, 400–408. doi:10.1359/jbmr.071103

Liu, S., Anwari, V., Yazdankhah, N., Tam, K., Whyte, R., Ha, E., et al. (2022). Bone-muscle-fat interactions and perfusive evidence of inflammation in non-overweight postmenopausal women with early knee oa. *Osteoarthr. Imaging* 2, 100030. doi:10.1016/j.ostima.2022.100030

Makrogiannis, S., Boukari, F., and Ferrucci, L. (2018). Automated skeletal tissue quantification in the lower leg using peripheral quantitative computed tomography. *Physiol. Meas.* 39, 035011. doi:10.1088/1361-6579/aaafb5

Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imaging* 22, 120–128. doi:10.1109/TMI.2003.809072

McAuliffe, M. J., Lalonde, F. M., McGarry, D., Gandler, W., Csaky, K., and Trus, B. L. (2001). Medical image processing, analysis and visualization in clinical research. IEEE symposium on computer-based medical systems CBMS. 26-27 July 2001. Bethesda, MD, USA. IEEE Computer Society, 381–386.

Owen, P. J., Hart, N. H., Latella, C., Hendy, A. M., Lamon, S., and Rantalainen, T. (2019). Identifying and assessing inter-muscular fat at the distal diaphyseal femur measured by peripheral quantitative computed tomography (pqct). *J. Clin. Densitom.* 24, 106–111. doi:10.1016/j.jocd.2019.11.001

Pauchard, Y., Ayres, F. J., and Boyd, S. K. (2011). Automated quantification of three-dimensional subject motion to monitor image quality in high-resolution peripheral quantitative computed tomography. *Phys. Med. Biol.* 56, 6523–6543. doi:10.1088/0031-9155/56/20/001

Pialat, J., Burghardt, A., Sode, M., Link, T., and Majumdar, S. (2012). Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. *Bone* 50, 111–118. doi:10.1016/j.bone.2011.10.003

Rodrigues, R., and Pinheiro, A. M. G. Segmentation of skeletal muscle in thigh dixon mri based on texture analysis. arXiv preprint arXiv:1904.04747 (2019).

Rogelj, P., and Kovačič, S. (2006). Symmetric image registration. *Med. Image Anal.* 10, 484–493. doi:10.1016/j.media.2005.03.003

Rohlfing, T., Brandt, R., Maurer, C. R., and Menzel, R. (2001). Bee brains, b-splines and computational democracy: Generating an average shape atlas. Proceedings IEEE workshop on mathematical methods in biomedical image analysis (MMBIA 2001). 09-10 December 2001. Kauai, HI, USA. IEEE, 187–194.

Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, J., and Calvin, R. (2005). *Quo vadis, atlas-based segmentation? The handbook of medical image analysis-volume iii: Registration models*.

Rowe, G. S., Blazevich, A. J., and Haff, G. G. (2019). pqct-and ultrasound-based muscle and fat estimate errors after resistance exercise. *Med. Sci. Sports Exerc.* 51, 1022–1031. doi:10.1249/MSS.0000000000001873

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imaging* 18, 712–721. doi:10.1109/42.796284

Shen, Dinggang, and Hammer, Davatzikos C. (2002). Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21, 1421–1439. doi:10.1109/TMI.2002.803111

Simon, A., Schäfer, H. S., Schmidt, F. N., Stürznickel, J., Amling, M., and Rolvien, T. (2022). Compartment-specific effects of muscle strength on bone microarchitecture in women at high risk of osteoporosis. *J. Cachexia. Sarcopenia and Muscle* 13, 2310–2321. doi:10.1002/jcsm.13044

Sode, M., Burghardt, A. J., Pialat, J. B., Link, T. M., and Majumdar, S. (2011). Quantitative characterization of subject motion in hr-pqct images of the distal radius and tibia. *Bone* 48, 1291–1297. doi:10.1016/j.bone.2011.03.755

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Trans. Med. Imaging* 32, 1153–1190. doi:10.1109/TMI.2013.2265603

Starr, J. F., Bandeira, L. C., Agarwal, S., Shah, A. M., Nishiyama, K. K., Hu, Y., et al. (2018). Robust trabecular microstructure in type 2 diabetes revealed by individual trabecula segmentation analysis of hr-pqct images. *J. Bone Min. Res.* 33, 1665–1675. doi:10.1002/jbmr.3465

Töpfer, D., Gerner, B., Finzel, S., Kraus, S., Museyko, O., Schett, G., et al. (2015). Automated three-dimensional registration of high-resolution peripheral quantitative computed tomography data to quantify size and shape changes of arthritic bone erosions. *Rheumatology* 54, 2171–2180. doi:10.1093/rheumatology/kev256

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2007). Non-parametric diffeomorphic image registration with the demons algorithm. International conference on medical image computing and computer-assisted intervention. Springer, 319–326.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008). Symmetric log-domain diffeomorphic registration: A demons-based approach. International conference on medical image computing and computer-assisted intervention. Springer, 754–761.

Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., and Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 611–623. doi:10.1109/TPAMI.2012.143

Warfield, S. K., Zou, K. H., and Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23, 903–921. doi:10.1109/TMI.2004.828354

Weston, A. D., Korfiatis, P., Kline, T. L., Philbrick, K. A., Kostandy, P., Sakinis, T., et al. (2019). Automated abdominal segmentation of CT scans for body composition analysis using deep learning. *Radiology* 290, 669–679. doi:10.1148/radiol.2018181432

Wong, A. K. (2016). A comparison of peripheral imaging technologies for bone and muscle quantification: A technical review of image acquisition. *J. Musculoskelet. Neuronal Interact.* 16, 265–282.

Wong, A. K. O., and Manske, S. L. (2020). A comparison of peripheral imaging technologies for bone and muscle quantification: A review of segmentation techniques. *J. Clin. Densitom.* 23, 92–107. doi:10.1016/j.jocd.2018.04.001

frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 23 November 2022
DOI 10.3389/fphys.2022.977427

Check for updates

OPEN ACCESS

EDITED BY
William C Cho,
QEH, Hong Kong SAR, China

REVIEWED BY
Hexiang Wang,
The Affiliated Hospital of Qingdao University, China
Zhan Feng,
Zhejiang University, China

*CORRESPONDENCE
Adili Abulaiti,
27984673@qq.com
Xiaoguang Zou,
zxgks@163.com

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

SPECIALTY SECTION
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

RECEIVED 05 July 2022
ACCEPTED 11 November 2022
PUBLISHED 23 November 2022

CITATION
Nijiati M, Tuersun A, Zhang Y, Yuan Q, Gong P, Abulizi A, Tuoheti A, Abulaiti A and Zou X (2022), A symmetric prior knowledge based deep learning model for intracerebral hemorrhage lesion segmentation.
Front. Physiol. 13:977427.
doi: 10.3389/fphys.2022.977427

COPYRIGHT
© 2022 Nijiati, Tuersun, Zhang, Yuan, Gong, Abulizi, Tuoheti, Abulaiti and Zou. 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 symmetric prior knowledge based deep learning model for intracerebral hemorrhage lesion segmentation

Mayidili Nijiati[1†], Abudouresuli Tuersun[1†], Yue Zhang[2†], Qing Yuan[1], Ping Gong[2], Abudoukeyoumujiang Abulizi[1], Awanisa Tuoheti[1], Adili Abulaiti[1*] and Xiaoguang Zou[3*]

[1]Department of Radiology, The First People's Hospital of Kashi Prefecture, Kashi, China, [2]Deepwise AI Lab, Beijing, China, [3]Clinical Medical Research Center, The First People's Hospital of Kashi Prefecture, Kashi, China

Background: Accurate localization and classification of intracerebral hemorrhage (ICH) lesions are of great significance for the treatment and prognosis of patients with ICH. The purpose of this study is to develop a symmetric prior knowledge based deep learning model to segment ICH lesions in computed tomography (CT).

Methods: A novel symmetric Transformer network (Sym-TransNet) is designed to segment ICH lesions in CT images. A cohort of 1,157 patients diagnosed with ICH is established to train (n = 857), validate (n = 100), and test (n = 200) the Sym-TransNet. A healthy cohort of 200 subjects is added, establishing a test set with balanced positive and negative cases (n = 400), to further evaluate the accuracy, sensitivity, and specificity of the diagnosis of ICH. The segmentation results are obtained after data pre-processing and Sym-TransNet. The DICE coefficient is used to evaluate the similarity between the segmentation results and the segmentation gold standard. Furthermore, some recent deep learning methods are reproduced to compare with Sym-TransNet, and statistical analysis is performed to prove the statistical significance of the proposed method. Ablation experiments are conducted to prove that each component in Sym-TransNet could effectively improve the DICE coefficient of ICH lesions.

Results: For the segmentation of ICH lesions, the DICE coefficient of Sym-TransNet is 0.716 ± 0.031 in the test set which contains 200 CT images of ICH. The DICE coefficients of five subtypes of ICH, including intraparenchymal hemorrhage (IPH), intraventricular hemorrhage (IVH), extradural hemorrhage (EDH), subdural hemorrhage (SDH), and subarachnoid hemorrhage (SAH), are 0.784 ± 0.039, 0.680 ± 0.049, 0.359 ± 0.186, 0.534 ± 0.455, and 0.337 ± 0.044, respectively. Statistical results show that the proposed Sym-TransNet can significantly improve the DICE coefficient of ICH lesions in most cases. In addition, the accuracy, sensitivity, and specificity of Sym-TransNet in the diagnosis of ICH in 400 CT images are 91.25%, 98.50%, and 84.00%, respectively.

Conclusion: Compared with recent mainstream deep learning methods, the proposed Sym-TransNet can segment and identify different types of lesions from CT images of ICH patients more effectively. Moreover, the Sym-TransNet can diagnose ICH more stably and efficiently, which has clinical application prospects.

KEYWORDS

intracerebral hemorrhage, lesion segmentation, deep learning, symmetric knowledge, transformer

Introduction

Intracerebral hemorrhage (ICH) is one of the most devastating subtypes of stroke, accounting for 10%–20% of all stroke cases (Zhao et al., 2020). ICH is commonly caused by trauma, hypertension, and vascular malformation, and more than half of the ICH patients have a long-term disabilities (Rindler et al., 2020). According to symptom onset time, the patients are divided into hyperacute stage (≤6 h), acute stage (7–72 h), subacute stage (3 days–2 weeks), and chronic stage (after 2 weeks) (Vijayan and Reddy, 2016). Depending on the hemorrhage site, ICH can be divided into five types, which include intraparenchymal hemorrhage (IPH), intraventricular hemorrhage (IVH), extradural hemorrhage (EDH), subdural hemorrhage (SDH), and subarachnoid hemorrhage (SAH) (Chilamkurthy et al., 2018). Different bleeding types determine the treatment plan and prognosis of patients. Therefore, accurate detection and classification of ICH are of great significance for saving the life and neurological function of patients (Li et al., 2021).

Neuroimaging is an important tool for the detection, characterization, and prediction of acute stroke, including ischemic and hemorrhagic subtypes (Lee et al., 2019). Computed tomography (CT) is the first choice for emergency diagnosis of ICH due to its high imaging speed (Chan, 2007). However, reading and analyzing a large amount of CT images is time-consuming and tricky work for clinic doctors, which increases the possibility of missed diagnosis and misdiagnosis (Cho et al., 2019). At present, emergency craniocerebral CT diagnosis, especially on the night shift, is mostly provided by the junior radiologist, and then reviewed by the senior radiologist (Lal et al., 2000; Erly et al., 2002). Several studies have shown that initial diagnosis provided by junior radiologists has different degrees of missed diagnosis and misdiagnosis (Erly et al., 2002). However, due to the high variability of the location, contrast, and shape of bleeds, accurate localization of them can be challenging and time-consuming even for experienced radiologists. In addition, due to limited medical conditions and resources in some underdeveloped areas, patients with ICH cannot receive an accurate diagnosis and timely treatment the first time, resulting in a threat to patients life. Therefore, it is very important to diagnose and classify ICH timely and accurate (Li et al., 2021).

Artificial intelligence (AI) technology is a rapidly developing field, which is regarded as a promising approach for fast and efficient image analysis (He et al., 2016; Li Y. et al., 2022). In recent years, AI has been applied in the medical imaging field of acute cerebrovascular diseases, including as a tool for classification, quantification, monitoring, and prediction (Ironside et al., 2019; Sun et al., 2020; Zhu et al., 2020). The convolutional neural network (CNN) is one of the representative deep learning algorithms that utilize image high-dimensional digital information by extracting image features (Badrinarayanan et al., 2017; Roy et al., 2019). In the field of medical image segmentation, the U-shaped network (U-Net) is one of the most representative convolutional neural networks (Ronneberger et al., 2015). In several years, many deep learning methods based on convolutional neural networks have been successfully applied to ICH lesion segmentation and achieved relatively ideal results. Inkeaw et al. proposed a 3D convolutional neural network, which processes CT images with different resolutions through four parallel paths, and segments different types of bleeding lesions through the region-growing method. The median DICE coefficient of segmentation for each bleeding subtype was higher than 0.37 (Inkeaw et al., 2022). Xu et al. (2021) adopted the densely connected U-Net architecture to test on nearly 300 ICH images and achieved a DICE coefficient of 0.89. Nevertheless, IVH and SAH were not included in this study. A supervised multi-task aiding representation transfer learning network (SMART-Net) was proposed to overcome the complex training process of the current deep learning model and the inefficient prediction accuracy on the patient's level (Kyung et al., 2022). Kwon et al. (2019) utilized a healthy brain template as auxiliary information for segmentation and employed U-Net to capture the difference between the input CT image and healthy template to segment ICH lesions more efficiently. The generative adversarial network (GAN) is also a common approach used in medical image segmentation tasks. A residual segmentation method with GAN (ReSGAN) was designed to learn a distribution of pseudo-normal brain CT scans and delineate the hemorrhaging areas (Toikkanen et al., 2021). To capture the interaction information between adjacent hematoma slices in CT images, Li et al. designed a slice expansion module and proposed two information transmission paths to expand the forward/backward slice respectively (Li X. Y. et al., 2022). The complicated annotation process of ICH lesions in CT images is

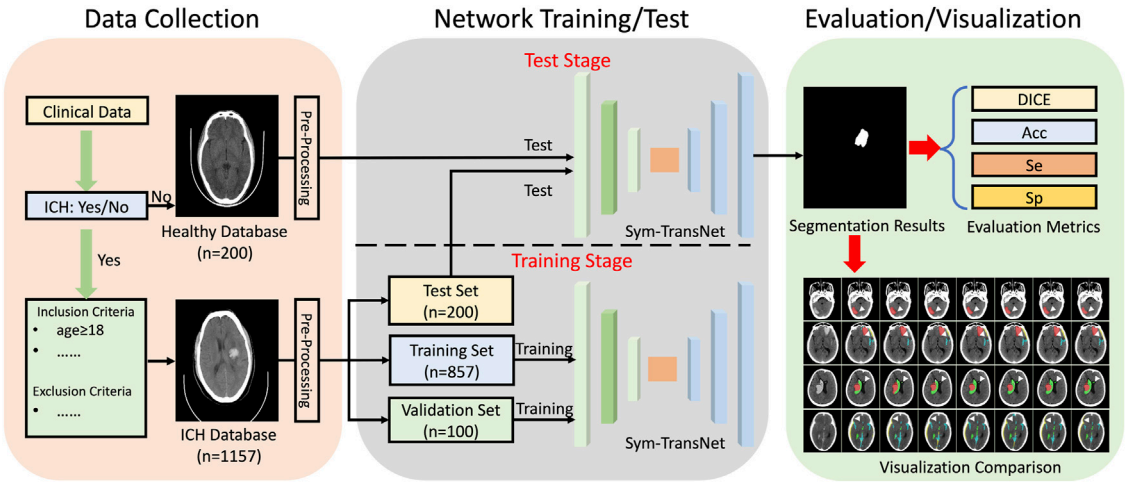


FIGURE 1
The workflow of the study.

one of the important factors which restricts the segmentation performance of the deep learning model. In order to make efficient utilization of unlabeled data, the semi-supervised learning approach such as the mean-teacher framework has also been transplanted by researchers for lesion segmentation (Cui et al., 2019).

However, the lesions of some hemorrhage types, such as SAH, are extremely extensive in the brain. The CNN-based methods which utilize local convolution kernel to obtain image features are difficult to effectively capture the long-distance dependencies in CT (Guo and Terzopoulos, 2021; Huang et al., 2022). Currently, the combination of the Transformer structure and CNN has been proven to be beneficial for capturing long-distance dependencies in images, which inspired us to use the Transformer structure for the segmentation of ICH lesions in this paper (Liu et al., 2020). In addition, the structure of the brain is roughly symmetrical (Liang et al., 2021). In a hemorrhagic stroke, this symmetry is commonly broken. Therefore, the symmetry change of brain structure can also be utilized for the segmentation of ICH lesions in CT images.

To solve the problem that the traditional CNN-based methods are difficult to capture the long-distance dependencies of CT images and the insufficient utilization of the symmetric structure of the brain, we proposed a novel deep learning method called symmetric Transformer network (Sym-TransNet) in this paper. The Sym-TransNet combines the Transformer structure with the traditional U-Net and adopts the symmetry prior knowledge in the network, which effectively improves the accuracy of segmentation and classification of ICH, reducing the workload of clinicians and providing a certain clinical basis for timely and accurate treatment of patients with ICH.

Materials and methods

In this retrospective study, a large number of CT images were collected clinically for training and testing of the proposed Sym-TransNet. Then we calculate the performance indicators of the proposed Sym-TransNet in the segmentation and diagnosis task of ICH, and visualize the results. The specific process is shown in Figure 1. Ethical approval for this study was waived by The Medical Ethics Committee of The First People's Hospital of Kashi Prefecture because this study used anonymous data which was collected as part of routine diagnosis and treatment.

Patients

A consecutive non-contrast head CT dataset, which retrospectively enrolled 1,157 patients who were diagnosed with ICH from January 2019 to April 2022 at the First People's Hospital of Kashi Prefecture, was established in this study for the training, validation, and test of our proposed deep learning model. We developed three patient inclusion criteria as follows: 1) patients (age≥18 years) who were diagnosed with ICH; 2) the diagnoses coincided with non-contrast head CT scans and radiology reports; 3) the CT scans were performed within 3 days after onset of symptoms. In addition, we also integrated three exclusion criteria: 1) patients who refused to sign informed consent; 2) CT Scans with excessive motion/artifacts (image quality not suitable for ICH diagnosis); 3) patients with both hemorrhagic and ischemic strokes. In addition, CT scans of 200 healthy subjects were collected to evaluate the diagnostic specificity of our deep learning model. These CT scans were obtained during the physical examination

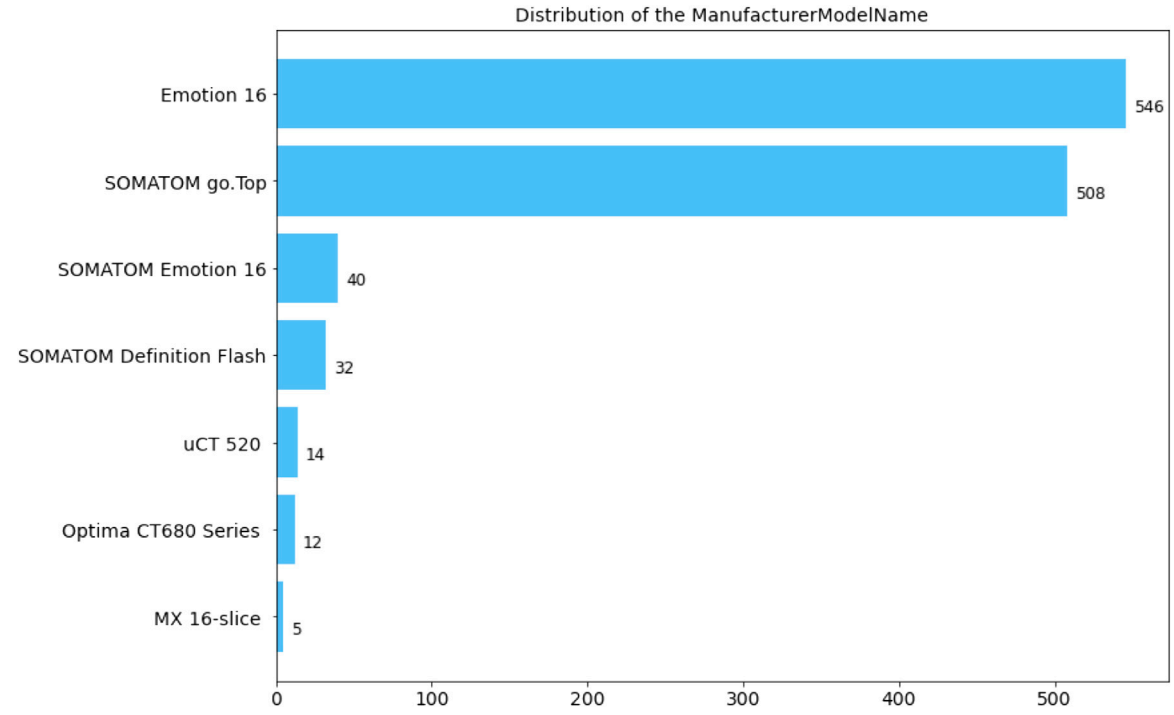


FIGURE 2
Distribution of CT information in the dataset. (a) Distribution of patient sex. O: unknown; F: female; M: male. (b) Distribution of the manufacturer model name.

in the First People's Hospital of Kashi Prefecture from September 2019 to April 2022.

Data collection and annotation

All CT scans diagnosed with ICH utilized in this research were obtained by CT scanners produced by Siemens, Philips, United Imaging, and General Medical System. The slice thickness of the CT scans is mainly 5 mm. Specifically, the numerical distribution of the manufacturer model name in our ICH dataset is displayed in Figure 2. The dataset contained 678 male and 478 female patients with intracerebral hemorrhage, and the gender of one scan was unknown. In our ICH dataset, CT slices of 857 patients are adopted to train the deep learning model, CT slices of 100 patients are used to adjust parameters during the training stage, and CT slices of 200 patients are employed to evaluate the overall segmentation performance. After anonymizing sensitive information in original DICOM (Digital Imaging and Communications in Medicine) data, lesions in five sub-types of ICH, including IPH, IVH, EDH, SDH, and SAH, were annotated by six experienced radiologists. Based on the above stages, two senior neurologists with more than 5 years of experience corrected the mislabeling of preliminary lesion annotations and further refined the outline, location, and categories of hemorrhage lesions. The final segmentation gold standard was determined by senior neurologists after reaching a consensus.

Data pre-processing

To better adapt to the training stage of the deep learning network, we pre-process the CT images in our dataset. Specifically, the window level (WL)/window width (WW) of each CT image is set as 50/100 and normalized to have zero mean and unit variance (Yu et al., 2022). To reduce the consumption of computing resources, the size of each CT slice is resized to 512 × 512. In the test stage, all segmentation results obtained by the proposed model are upsampled to the original size for performance index calculation. Furthermore, some CT slices in the dataset are randomly flipped horizontally to increase the diversity of the training data. Due to the head position of the patient during the CT scan is not uniquely deterministic, the reconstructed brain structure in the CT slice is usually not horizontally symmetrical. To effectively utilize the medical prior knowledge that the brain structure is symmetrical, the symmetry-based alignment network (Wang et al., 2020) is utilized in this study to horizontally align the brain in all CT slices. As shown in Figure 3, the brain structure in the CT slice is transformed from asymmetric to horizontal symmetry after being processed by the alignment network.

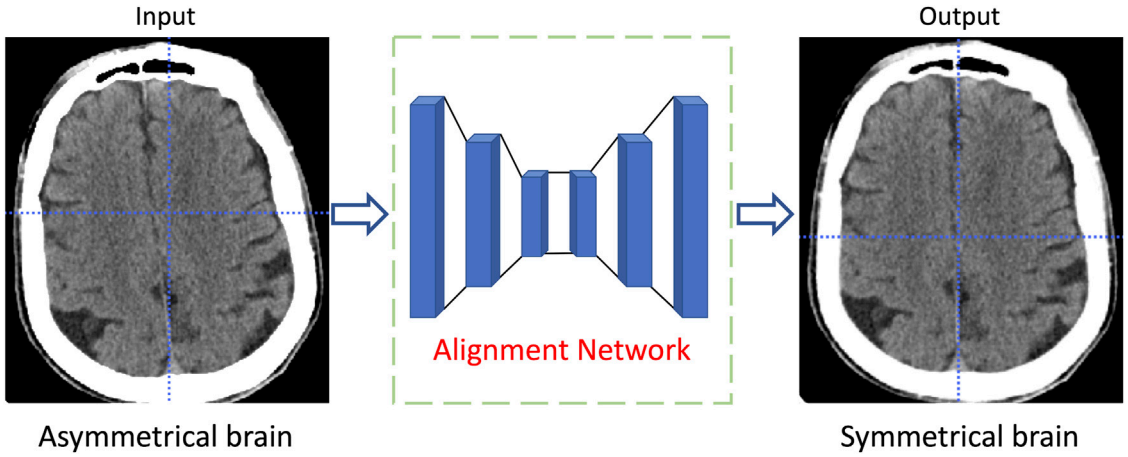


FIGURE 3
The diagram of the CT slice alignment through symmetric based alignment network.

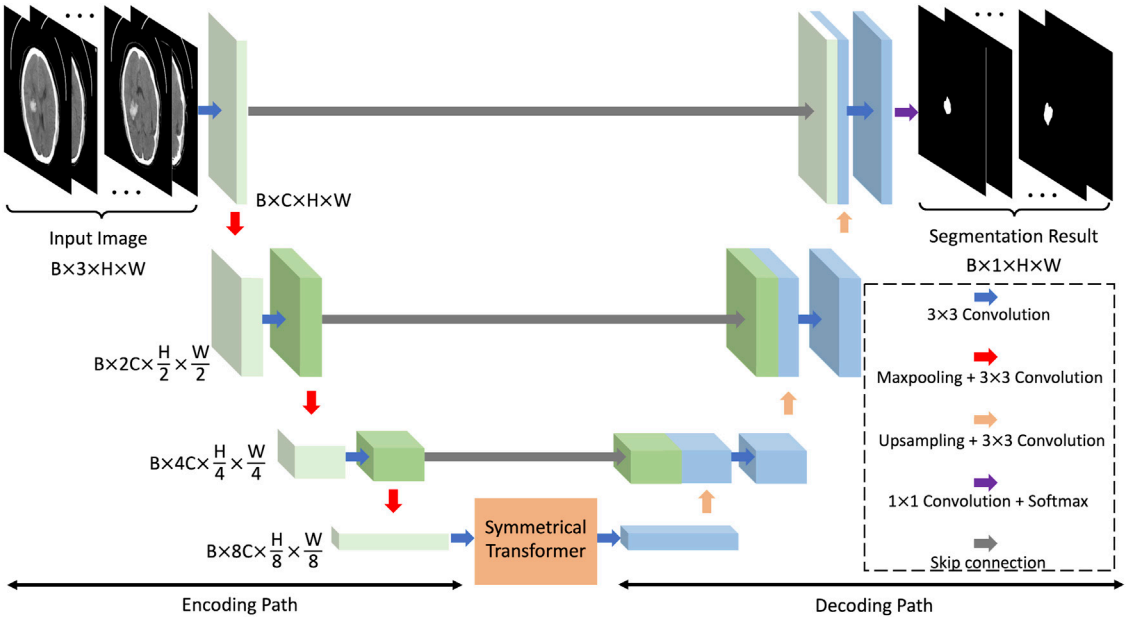


FIGURE 4
The framework of the proposed Sym-TransNet.

Symmetric prior knowledge-based deep learning model

The convolutional neural network, which obtains high-level features in images through the convolution kernels and the down-sampling operation, is an efficient deep learning model utilized in various research fields. In the field of medical image segmentation, the U-Net is one of the most representative convolutional neural networks. The U-Net extracts and restores image features through interconnected codec paths, and has satisfactory performance in different segmentation tasks (Ibtehaz and Rahman, 2020). As shown in Figure 4, the proposed deep learning model is based on the U-Net framework, containing an image encoding path and an image decoding path.

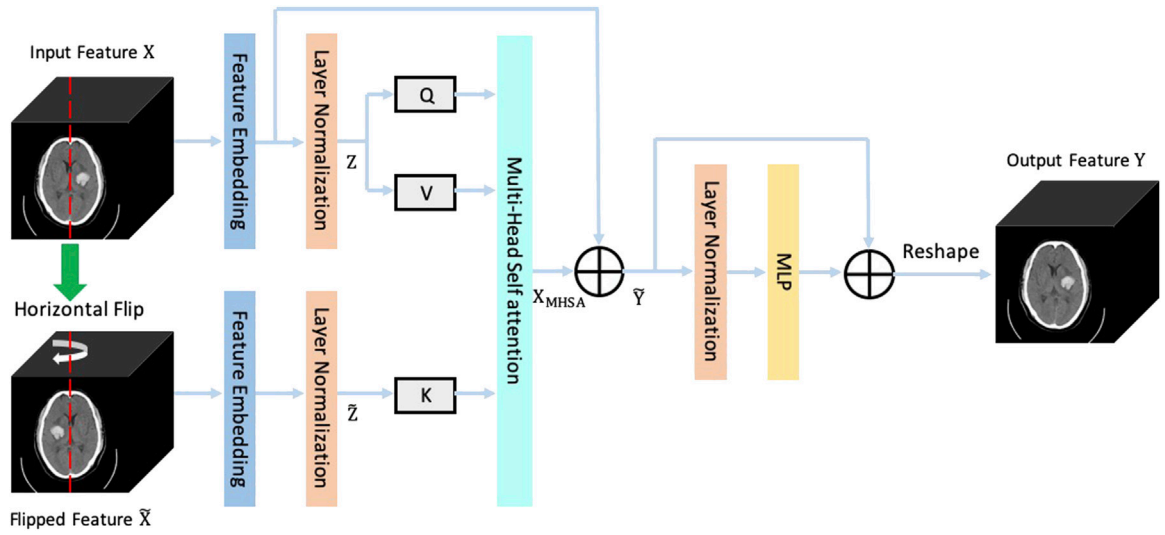


FIGURE 5
The diagram of the symmetrical Transformer.

The bottom of the two paths is connected by a symmetric Transformer. We name this network the symmetric Transformer network (Sym-TransNet). The image preprocessed by the alignment network, which is regarded as the input of our Sym-TransNet, is transformed into a tensor of size $B \times C \times H \times W$ by the 3×3 convolution kernel, where B is the batch size which is pre-set in the training stage, C is the number of image channels, and H and W are the height and width of the image. In this paper, parameters B, C, H, and W are set to 64, 64, 512, and 512, respectively. The encoding path captures the high-level semantic information in CT images by using continuous 3×3 convolution and max-pooling and finally generates image features with the size of $B \times 8C \times \frac{H}{8} \times \frac{W}{8}$.

The structure of the brain is roughly symmetrical. When a hemorrhagic stroke occurs, this symmetry is destroyed. Therefore, we can segment the bleeding area by capturing changes in symmetry in the brain CT image. However, the CNN which employs local convolution kernels is difficult to efficiently capture the long-distance symmetric relations in CT images. In recent years, the Transformer structure becomes a powerful approach for capturing long-distance dependencies in images and has achieved great success in computer vision tasks. Based on the medical prior knowledge that the brain structure is symmetrical, we propose a new Transformer structure named symmetrical Transformer to model long-distance symmetric relations in brain CT images. The structure of the proposed symmetrical Transformer is shown in Figure 5. We regard the image feature X obtained from the coding path as the input of the symmetric Transformer, and then flip it horizontally to get the flipped feature $\tilde{X}$. The feature embedding and the layer normalization are employed to map the above two features into sequence features Z and $\tilde{Z}$, which are both adapted to the Transformer structure. From Figure 5, the symmetrical Transformer mainly consists of two parts: the multi-head self-attention $\mathrm{MHSA}(\cdot)$ and multi-layer perceptron $\mathrm{MLP}(\cdot)$. The input Q, V, and K of the MHSA can be calculated as follows:

$$Q = ZW^Q \tag{1}$$

$$V = ZW^V \tag{2}$$

$$K = \tilde{Z}W^K \tag{3}$$

where the W^Q, W^V, and W^K are learnable parameters, called a query transform matrix, a value transform matrix, and a key transform matrix, respectively. Further, the output of the MHSA X_{MHSA} can be calculated as:

$$X_{\mathrm{MHSA}} = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{4}$$

where d is a scaling factor that can solve the small gradient of the Softmax function. Thus, the output of the symmetrical Transformer Y can be represented as follows:

$$Y = \mathrm{Reshape}\left(\mathrm{MLP}\left(\mathrm{LN}\left(\tilde{Y}\right)\right) + \tilde{Y}\right) \tag{5}$$

$$\tilde{Y} = X_{\mathrm{MHSA}} + \mathrm{Feature\ Embedding}(X) \tag{6}$$

where $\mathrm{Reshape}(\cdot)$, $\mathrm{MLP}(\cdot)$, $\mathrm{LN}(\cdot)$, and $\mathrm{Feature\ Embedding}(\cdot)$ denote image reshape operation, multi-layer perceptron, layer normalization, and feature embedding operation.

The features processed by the symmetrical Transformer are fed to the decoding path, and the image features are restored to the original resolution through continuous convolution operation and upsampling operation. Similar to the traditional

U-Net, we adopt a skip connection between the encoding path and decoding path to obtain multi-scale features. Finally, 1 × 1 convolution and softmax function are utilized to generate the final segmentation result.

Model training and evaluation metrics

The proposed Sym-TransNet is implemented based on PyTorch deep learning framework and is trained and evaluated on 4 NVIDIA RTX graphic cards. In the training stage, the Adam optimizer is adopted to optimize model parameters. In addition, cross-entropy and DICE loss functions are used to measure the distance between the segmentation results of the deep learning model and the golden standards during gradient backpropagation. We set the initial learning rate as 1×10^{-5} and the parameter weight decay as 1×10^{-3}, then utilize the exponential decay strategy to dynamically adjust the learning rate to avoid the local minimum. We set the maximum training epoch as 100, selecting the model with the best performance in the validation set to conduct performance evaluation on the test set.

In the test stage, the DICE coefficient is employed to evaluate the accuracy of the segmentation of ICH lesions. Assuming P is the prediction result of our proposed Sym-TransNet and G is the segmentation gold standard, the DICE coefficient can be calculated as:

$$DICE = \frac{2 \times |P \cap G|}{|P \cup G|} \tag{7}$$

In addition to calculating the DICE coefficient of lesion segmentation, we also evaluate the accuracy (Acc), sensitivity (Sen), and specificity (Spe) of the proposed Sym-TransNet on the ICH diagnosis task. The diagnosis results are divided into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Thus, the Acc, Sen, and Spe can be calculated as follow:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

$$Sen = \frac{TP}{TP + FN} \tag{9}$$

$$Spe = \frac{TN}{TN + FP} \tag{10}$$

Statistical analysis

To statistically analyze the segmentation results obtained by the deep learning model, we utilized Wilcox rank-sum test to conduct pair-wise statistical tests (on DICE coefficient) between the proposed Sym-TransNet and several existing deep learning methods which are widely used in medical imaging (Wu et al., 2021). All of the statistical analysis in this paper was implemented in Python. We defined that the two methods were statistically different when the p-value <0.05.

Results

Satisfactory performance of Sym-TransNet for ICH lesion segmentation

After the parameters of the proposed Sym-TransNet are optimized on the training dataset, we evaluate the performance on the test dataset. To demonstrate that the proposed method has satisfactory segmentation performance compared with existing deep learning models, we faithfully reproduce several approaches commonly utilized in the field of medical image processing in recent years for comparison, including the U-Net, the U-Net with the dilated convolution (DU-Net) (Yu et al., 2017), the U-Net with SE block (SEU-Net) (Roy et al., 2019), the Dual-Attention Network (DA-Net) (Fu et al., 2019), and the High-Resolution Network (HR-Net) (Sun et al., 2019; Wang et al., 2021). The performance of the above methods in the segmentation of ICH lesions and the segmentation of five subtypes of ICH lesions is listed in Table 1. As shown in Table 1, our Sym-TransNet achieves an average DICE of 0.716, where the 95% confidence interval (95% CI) is 0.685–0.747, on the test dataset containing 200 patients with ICH, which is the best performance compared with the current methods in Table 1. Furthermore, for the segmentation of different subtypes of ICH lesions, the average DICE of IPH, IVH, EDH, SDH, and SAH by the Sym-TransNet is 0.784 (95% CI: 0.745–0.824), 0.680 (95% CI: 0.631–0.730), 0.359 (95% CI: 0.173–0.545), 0.534 (95% CI: 0.455–0.613), and 0.337 (95% CI: 0.293–0.382). In terms of IPH segmentation, the HR-Net is the model with the highest DICE among the comparison methods in Table 1, and our Sym-TransNet has improved by 0.26 on this basis. In the IVH case, the Sym-TransNet reaches the highest DICE score of all methods and 0.26 higher than the second-place method HR-Net. In addition, compared with other methods, the segmentation DICE coefficient of our method in EDH and SDH has been significantly improved. The reason why the performance of SDH and EDH is not as satisfactory as IPH and IVH are that the two kinds of bleeding lesions are close to the skull, resulting in the symmetry prior knowledge is not significantly beneficial to distinguishing the two kinds of lesions when they appear at the same time. As the lesions of SAH are very irregular in shape and often diffuse into the sulci, the traditional CNN methods based on local information modeling have low segmentation performance for SAH. The Sym-TransNet has improved the segmentation of SAH compared with the method based on CNN alone because of the combination of the Transformer structure that can capture

TABLE 1 Segmentation performance of deep learning methods on ICH dataset.

Method	Parameters (1×10^6)	ICH lesions (95% CI)	IPH (95% CI)	IVH (95% CI)	EDH (95% CI)	SDH (95% CI)	SAH (95% CI)
U-Net	2.47	0.624 (0.587, 0.661)	0.688 (0.638, 0.738)	0.518 (0.457, 0.580)	0.222 (0.083, 0.361)	0.321 (0.239, 0.404)	0.245 (0.206, 0.284)
DU-Net	4.83	0.611 (0.573, 0.649)	0.674 (0.622, 0.725)	0.496 (0.432, 0.560)	0.119 (0.014, 0.224)	0.256 (0.171, 0.341)	0.226 (0.186, 0.265)
SEU-Net	2.47	0.629 (0.592, 0.666)	0.691 (0.642, 0.741)	0.538 (0.478, 0.598)	0.253 (0.101, 0.405)	0.381 (0.304, 0.459)	0.256 (0.216, 0.295)
DA-Net	26.97	0.669 (0.636, 0.702)	0.739 (0.695, 0.782)	0.605 (0.550, 0.660)	0.244 (0.112, 0.375)	0.434 (0.348, 0.520)	0.269 (0.227, 0.311)
HR-Net	17.12	0.686 (0.656, 0.717)	0.758 (0.715, 0.802)	0.654 (0.603, 0.706)	0.205 (0.046, 0.364)	0.531 (0.458, 0.606)	0.317 (0.306, 0.389)
Ours	43.06	0.716 (0.685, 0.747)	0.784 (0.745, 0.824)	0.680 (0.631, 0.730)	0.359 (0.173, 0.545)	0.534 (0.455, 0.613)	0.337 (0.293,0.382)

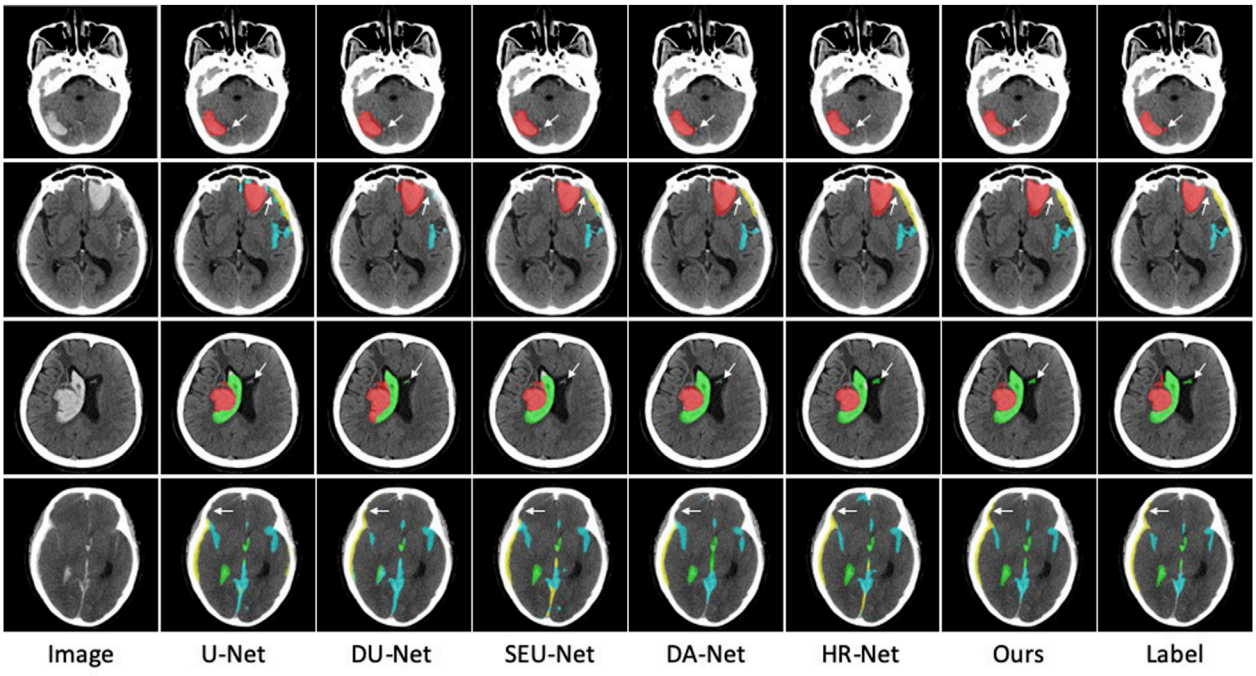


FIGURE 6
Visual comparison of segmentation results. The red, green, yellow, and blue pixels represent lesions of IPH, IVH, SDH, and SAH.

the long-distance dependencies in the CT scan. The HR-Net is the best method among the compared methods for SAH segmentation because the multi-resolution method utilized in HR-Net captures more global information in SAH than other CNN-based methods. Compared with the HR-Net, the proposed Sym-TransNet further improves the DICE coefficient of 0.2 on SAH, indicating that the Transformer structure is effective in the segmentation of SAH.

In addition to the DICE coefficient comparison, to more intuitively demonstrate the effectiveness of our method, the segmentation results of some ICH are also visualized in Figure 6. As far as the segmentation of cerebral hemorrhage lesions is concerned, the proposed Sym-TransNet can obtain the edges of lesions more consistent with manual labeling and can detect some small lesions missed by other methods. For IPH and IVH segmentation, when parenchymal hemorrhage breaks into

TABLE 2 Statistical test results (*p*-value) between Sym-TransNet and other methods.

Methods types	ICH lesions	IPH	IVH	EDH	SDH	SAH
U-Net	2.95×10^{-5}	1.09×10^{-5}	4.63×10^{-5}	$2.6\ 0 \times 10^{-1}$	8.20×10^{-4}	3.89×10^{-3}
DU-Net	7.06×10^{-6}	2.84×10^{-7}	9.73×10^{-6}	1.19×10^{-1}	9.25×10^{-5}	5.19×10^{-4}
SEU-Net	9.71×10^{-5}	3.69×10^{-5}	2.33×10^{-4}	2.25×10^{-1}	5.86×10^{-3}	9.30×10^{-3}
DA-Net	8.22×10^{-3}	4.01×10^{-4}	1.62×10^{-2}	2.36×10^{-1}	9.08×10^{-2}	2.61×10^{-2}
HR-Net	4.04×10^{-2}	1.17×10^{-1}	3.41×10^{-1}	3.55×10^{-1}	8.29×10^{-1}	7.72×10^{-1}

TABLE 3 Ablation results of Sym-TransNet.

Method	ICH lesions (95% CI)	IPH (95%CI)	IVH (95%CI)	EDH (95%CI)	SDH (95%CI)	SAH (95%CI)
U-Net☑						
Transformer☒	0.624 (0.587, 0.661)	0.688 (0.638, 0.738)	0.518 (0.457, 0.580)	0.222 (0.083, 0.361)	0.321 (0.239, 0.404)	0.245 (0.206, 0.284)
Symmetric Prior☒						
U-Net☑						
Transformer☑	0.691 (0.659, 0.723)	0.761 (0.719, 0.802)	0.624 (0.569, 0.678)	0.233 (0.086, 0.379)	0.451 (0.370, 0.533)	0.284 (0.242, 0.326)
Symmetric Prior☒						
U-Net☑						
Transformer☑	0.716 (0.685, 0.747)	0.784 (0.745, 0.824)	0.680 (0.631, 0.730)	0.359 (0.173, 0.545)	0.534 (0.455, 0.613)	0.337 (0.293, 0.382)
Symmetric Prior☑						

the ventricle, the Sym-TransNet can more clearly define the interface between the two types of hemorrhage at the ventricle. For EDH and SDH, the Sym-TransNet reduces the risk of misidentifying the skull as bleeding. In addition, compared with other methods, our model can detect SAH diffusing in sulci more sensitively.

We also conduct pairwise statistical tests (between the Sym-TransNet and other methods) on the segmentation results of ICH lesions and five hemorrhage subtypes, as shown in Table 2. Except for EDH, the Sym-TransNet improves the segmentation performance of the U-Net, DU-Net, and SEU-Net with statistical significance (all p-values < 0.05). For DA-Net and Sym-TransNet, p-values were all less than 0.05 in ICH lesion, IPH, IVH, and SAH. For HR-Net, our Sym-TransNet improved ICH lesion segmentation significantly, but there was no statistical difference in the segmentation of the five bleeding subtypes.

Transformer and symmetric prior knowledge improves the segmentation accuracy

To prove the validity of the Transformer structure and symmetric prior knowledge utilized in this paper, we conduct ablation experiments for the Transformer structure and symmetric prior knowledge. The results of the ablation experiment are shown in Table 3. In the ablation experiment, the original U-Net is regarded as the baseline method, and the Transformer structure and symmetric prior knowledge are combined to observe the changes in the segmentation performance. As shown in Table 3, with the addition of the Transformer structure, the DICE coefficient of ICH lesions is significantly improved (from 0.624 to 0.691). In addition, the DICE coefficient of five different subtypes of ICH lesions is also increased (IPH: from 0.688 to 0.761, IVH: from 0.518 to 0.624, EDH: from 0.222 to 0.233, SDH: from 0.321 to 0.451, SAH: from 0.245 to 0.284). The improved segmentation performance of SAH lesions indicates that the combination of U-Net and Transformer structure is beneficial to the network to capture long-distance dependence in CT images and improve the segmentation accuracy of irregular lesions. Furthermore, after the fusion of symmetric prior knowledge, the segmentation performance of ICH lesions is improved (from 0.691 to 0.716). The segmentation DICE of the five subtypes also increase to different degrees (IPH increased by 0.023, IVH increased by 0.056, EDH increased by 0.126, SDH increased by 0.083, SAH increased by 0.053). The most obvious improvement is in the segmentation of IPH lesions because IPH contains abundant symmetric information (IPH lesions usually only appear on one side of the brain tissue), which

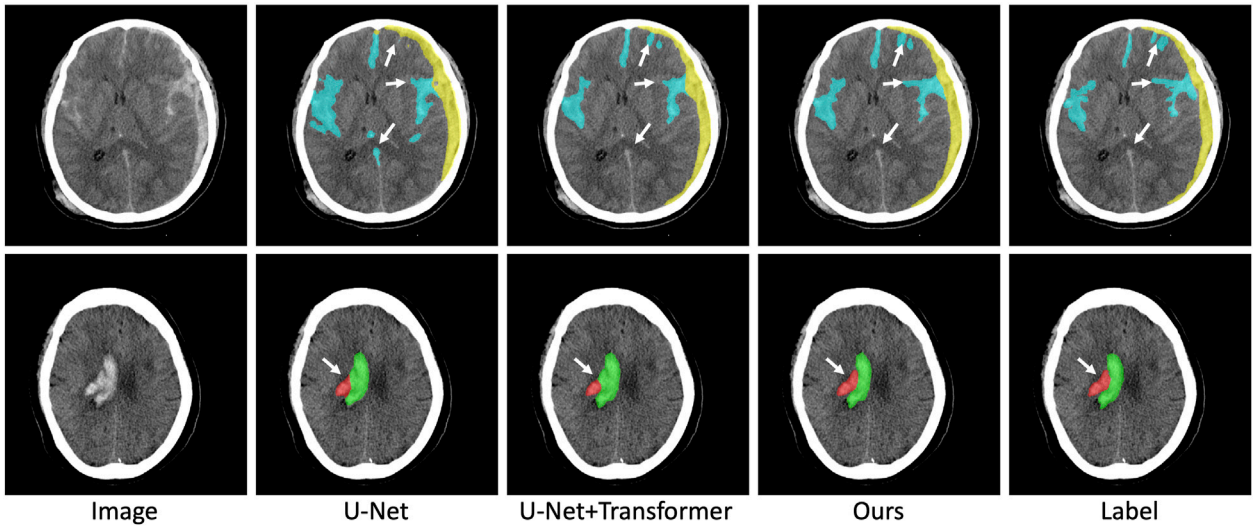


FIGURE 7
Visualization results of ablation experiments. The red, green, yellow, and blue pixels represent lesions of IPH, IVH, SDH, and SAH.

indicates that the combination of symmetric information in the network is conducive to the improvement of the segmentation accuracy of ICH lesions. We visualize the results of the ablation experiment in Figure 7. It can be seen from Figure 7 that the combination of Transformer structure with U-Net results in clearer edge details of lesions and reduced part of false positive lesions. In particular, the Transformer structure improves the segmentation result of the deep learning model for SAH lesions which are widely distributed in the CT image, and reduces false positives in U-Net segmentation results, demonstrating that the Transformer structure can effectively capture the long-distance dependencies in the CT image. In addition, with the utilization of symmetric prior knowledge, the segmentation results of IPH and IVH are closer to the gold standard than the other two methods. As shown in Figure 7, the use of symmetric prior knowledge also refines the interface between IPH and IVH lesions, further improving the segmentation performance of the model.

Sym-TransNet effectively diagnose patients with ICH

Clinically, the accurate diagnosis of ICH is of great significance for the follow-up treatment of patients. To demonstrate that the proposed deep learning model can sensitively detect patients with ICH while maintaining considerable specificity, we analyze the accuracy, sensitivity, and specificity of our model on 400 CT data (200 positive cases and 200 negative cases). Specifically, after the model obtains the segmentation results of the above 400 CT images, we determine whether the model classifies the corresponding data as ICH by calculating whether there are lesions in the segmentation results. Assuming that the data is a positive case and the segmentation result includes lesions, we record the segmentation result of this case as true positive (TP), and *vice versa* as false negative (FN). Assuming that the data is negative cases and there is no segmented lesion in the output of our model, we record the segmentation result of this case as true negative (TN), and otherwise as false positive (FP). As shown in Table 4, for the 200 positive samples, 197 cases are correctly diagnosed and only 3 cases are misdiagnosed as negative. Of the 200 negative samples, 168 are identified as negative and 32 were misdiagnosed as positive. Therefore, according to Eqs 8–10, we can calculate that the diagnostic accuracy, sensitivity, and specificity of the proposed model are 91.25%, 98.5%, and 84%, respectively. We also evaluate the diagnostic performance of the baseline method U-Net for ICH. The comparison of performance indexes between the two methods is shown in Table 5. Compared with the baseline method U-Net, the proposed Sym-TransNet has improved accuracy, sensitivity, and specificity (3%, 0.5%, and 5.5%, respectively), indicating that our method can effectively improve the overall performance of the ICH detection task based on the baseline model.

Discussion

In this study, we verify that the proposed Sym-TransNet has better segmentation performance than the existing mainstream deep learning methods in the segmentation of ICH lesions and the segmentation of five ICH subtypes. Compared with the

TABLE 4 Diagnostic results on CT data of 400 cases.

Total cases: 400 (positive: 200 negative: 200)	Positive (predict)	Negative (predict)
Positive (Actual)	197 (TP)	3 (FP)
Negative (Actual)	32 (FN)	168 (TN)

TABLE 5 Comparison of diagnostic performance of ICH.

Methods	Acc (%)	Sen (%)	Spe (%)
U-Net	88.25	98.00	78.50
Our Method	91.25	98.50	84.00

baseline model U-Net, the proposed deep learning model improves accuracy, sensitivity, and specificity in diagnosing ICH. Specifically, aiming at the five subtypes of ICH, the proposed Sym-TransNet achieves the highest DICE coefficient in intracerebral hemorrhage (comparison of DICE coefficients of five subtypes: IPH>IVH>SDH>EDH>SAH). Subdural hemorrhage and subarachnoid hemorrhage are the most easily missed or misdiagnosed subtypes of ICH due to irregular shape, unclear edge, and certain particularity of the site of hemorrhage. Subarachnoid hemorrhage is an example in which the lesions are filled with sulci, fissures, and cistern in casting shape, and the distribution is extensive and irregular. In addition, in the annotation process of segmentation gold standard, due to the existence of the CT volume effect, radiologists are difficult to accurately locate lesions of the subject who was diagnosed with EDH or SAH with less bleeding. The imprecise segmentation gold standard makes it difficult to train the deep learning model and directly affects its segmentation performance on the test set (the segmentation DICE coefficients of the model on EDH and SAH are 0.359 and 0.337, respectively). In order to alleviate the problems that exist in the process of annotation, many semi-supervised methods were proposed and have achieved better segmentation performance than fully supervised methods with a limited amount of labeled data. Therefore, the combination of the semi-supervised learning strategy may be an effective means to solve this problem, which is also our future research direction. In addition, researchers can also use the average value of multiple professional physicians as the final gold standard to alleviate the impact of inaccurate labeling on segmentation performance.

The segmentation performance of Sym-TransNet on IPH and IVH is significantly improved compared with existing methods. For the above two types of ICH, Sym-TransNet improved the DICE coefficient by 0.26 on the basis of the HR-Net. In terms of visualization of segmentation results, the segmentation results obtained by Sym-TransNet are more consistent with manual annotation, and some microscopic lesions missed by the other methods are detected. When IPH invades the ventricle, Sym-TransNet can more clearly distinguish the interface between the two types of hemorrhage. For EDH and SDH, Sym-TransNet reduced the risk of misidentifying the skull as a bleeding point. Moreover, compared with other deep learning methods, our model is more sensitive to detecting the SAH lesions in the sulci. Therefore, the proposed Sym-TransNet can more effectively segment different types of ICH lesions from CT images of ICH patients, which has potential clinical application prospects.

However, this study also has some limitations. From the perspective of data collection, as a retrospective study, selection bias may exist in this paper. Although a large number of CT data were included as the training set in this study, the sample size of the test set was insufficient, which probably results in accidental segmentation performance. Therefore, it is necessary to increase the test sample size to verify the model performance. Additionally, all of our CT scans came from a single center, and the diversity of samples can be further improved. From the perspective of the number of trainable parameters for deep learning models, the proposed Sym-TransNet is not optimal. We list the number of parameters for several mainstream models used for performance comparisons in this paper in Table 1. As shown in Table 1, Sym-TransNet has the highest number of model parameters. This is because the self-attention mechanism in the Transformer model requires a large amount of computation to obtain the long-distance dependence information. In the future, we will explore a more lightweight Transformer model for ICH lesion segmentation, which will be better applied in clinical practice.

In summary, Sym-TransNet proposed in this paper can accurately segment the ICH lesions and the five hemorrhage subtypes, improving the performance on the basis of the U-Net for the diagnosis of ICH. Sym-TransNet is expected to help relieve the workload of radiologists and reduce the rate of misdiagnosis of ICH in clinical practice, providing a basis for assisting clinical decision-making.

Data availability statement

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

Ethics statement

Ethical approval for this study was waived by the Medical Ethics Committee of The First People's Hospital of Kashi Prefecture because this study used anonymous data which was collected as part of routine diagnosis and treatment.

Author contributions

Study design and conception were proposed by MN and XZ; Paper writing was done by MN, A. Tuersun, YZ, PG, A. Abulizi, and XZ; ICH data was collected by MN, A. Tuersun, QY, A. Abulizi, A. Tuoheti, A. Abulaiti, and XZ; AI model training, testing and visualizing were done by YZ and PG. All authors interpreted the results and revised the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the Innovation Team Project of "Pearl River Scholar Tianshan Talents" Cooperative Expert Studio of The First People's Hospital of Kashi Prefecture (No. KDYY202017, MN) and "The application of standardized head and neck CTA scanning technique in early diagnosis of acute ischemic stroke in primary hospitals" (No. KSCG202208, MN).

Conflict of interest

Authors YZ and PG were employed by the company Deepwise AI Lab.

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.

References

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi:10.1109/TPAMI.2016.2644615

Chan, T. (2007). Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain. *Comput. Med. Imaging Graph.* 31, 285–298. doi:10.1016/j.compmedimag.2007.02.010

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., et al. (2018). Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* 392, 2388–2396. doi:10.1016/S0140-6736(18)31645-3

Cho, J., Park, K. S., Karki, M., Lee, E., Ko, S., Kim, J. K., et al. (2019). Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models. *J. Digit. Imaging* 32, 450–461. doi:10.1007/s10278-018-00172-1

Cui, W. H., Liu, Y. L., Li, Y. X., Guo, M. H., Li, Y. M., Li, X. L., et al. (2019). "Semi-supervised brain lesion segmentation with an adapted mean teacher model," Information Processing in Medical Imaging, 554–565.

Erly, W. K., Berger, W. G., Krupinski, E., Seeger, J. F., and Guisto, J. A. (2002). Radiology resident evaluation of head CT scan orders in the emergency department. *AJNR. Am. J. Neuroradiol.* 23, 103–107.

Fu, J., Liu, J., Tian, H. J., Li, Y., Bao, Y. J., Fang, Z. W., et al. (2019). "Dual attention network for scene segmentation," IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3141–3149.

Guo, D., and Terzopoulos, D. (2021). "A transformer-based network for anisotropic 3D medical image segmentation," IEEE International Conference on Patern Recognition, 8857–8861.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

Huang, S., Li, J., Xiao, Y., Shen, N., and Xu, T. (2022). RTNet: Relation transformer network for diabetic retinopathy multi-lesion segmentation. *IEEE Trans. Med. Imaging* 41, 1596–1607. doi:10.1109/TMI.2022.3143833

Ibtehaz, N., and Rahman, M. S. (2020). MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* 121, 74–87. doi:10.1016/j.neunet.2019.08.025

Inkeaw, P., Angkurawaranon, S., Khumrin, P., Inmutto, N., Traisathit, P., Chaijaruwanich, J., et al. (2022). Automatic hemorrhage segmentation on head CT scan for traumatic brain injury using 3D deep learning model. *Comput. Biol. Med.* 146, 105530. doi:10.1016/j.compbiomed.2022.105530

Ironside, N., Chen, C. J., Mutasa, S., Sim, J. L., Marfatia, S., Roh, D., et al. (2019). Fully automated segmentation algorithm for hematoma volumetric analysis in spontaneous intracerebral hemorrhage. *Stroke* 50, 3416–3423. doi:10.1161/STROKEAHA.119.026561

Kwon, D., Ahn, J., Kim, J., Choi, I., Jeong, S., Lee, Y. S., et al. (2019). "Siamese U-net with healthy template for accurate segmentation of intracranial hemorrhage," Medical Image Computing Computer Assisted Intervention, 848–855.

Kyung, S., Shin, K., Jeong, H., Kim, K. D., Park, J., Cho, K., et al. (2022). Improved performance and robustness of multi-task representation learning with consistency loss between pretexts for intracranial hemorrhage identification in head CT. *Med. Image Anal.* 81, 102489. doi:10.1016/j.media.2022.102489

Lal, N. R., Murray, U. M., Eldevik, O. P., and Desmond, J. S. (2000). Clinical consequences of misinterpretations of neuroradiologic CT scans by on-call radiology residents. *AJNR. Am. J. Neuroradiol.* 21, 124–129.

Lee, H., Yune, S., Mansouri, M., Kim, M., Tajmir, S. H., Guerrieri, C. E., et al. (2019). An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat. Biomed. Eng.* 3, 173–182. doi:10.1038/s41551-018-0324-9

Li, L., Wei, M., Liu, B., Atchaneeyasakul, K., Zhou, F. G., Pan, Z. H., et al. (2021). Deep learning for hemorrhagic lesion detection and segmentation on brain CT images. *IEEE J. Biomed. Health Inf.* 25, 1646–1659. doi:10.1109/JBHI.2020.3028243

Li, X. Y., Luo, G. N., Wang, W., Wang, K. Q., Gao, Y., and Li, S. (2022a). Hematoma expansion context guided intracranial hemorrhage segmentation and uncertainty estimation. *IEEE J. Biomed. Health Inf.* 26, 1140–1151. doi:10.1109/JBHI.2021.3103850

Li, Y., Zhang, Y., Cui, W., Lei, B., Kuang, X., and Zhang, T. (2022b). Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation. *IEEE Trans. Med. Imaging* 41, 1975–1989. doi:10.1109/TMI.2022.3151666

Liang, K. M., Han, K., Li, X. L., Cheng, X. Q., Li, Y. M., Wang, Y. Z., et al. (2021). "Symmetry-enhanced attention network for acute ischemic infarct segmentation

with non-contrast CT images," Medical Image Computing and Computer Assisted Intervention, 432–441.

Liu, Y., Yuan, H., Wang, Z., and Ji, S. (2020). Global pixel transformers for virtual staining of microscopy images. *IEEE Trans. Med. Imaging* 39, 2256–2266. doi:10.1109/TMI.2020.2968504

Rindler, R. S., Allen, J. W., Barrow, J. W., Pradilla, G., and Barrow, D. L. (2020). Neuroimaging of intracerebral hemorrhage. *Neurosurgery* 86, E414-E423–E423. doi:10.1093/neuros/nyaa029

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," Medical Image Computing Computer Assisted Intervention, 234–241.

Roy, A. G., Navab, N., and Wachinger, C. (2019). Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. *IEEE Trans. Med. Imaging* 38, 540–549. doi:10.1109/TMI.2018.2867261

Sun, K., Xiao, B., Liu, D., Wang, J. D., and Soc, I. C. (2019). "Deep high-resolution representation learning for human pose estimation," IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5686–5696.

Sun, L., Shao, W., Zhang, D., and Liu, M. (2020). Anatomical attention guided deep networks for ROI segmentation of brain MR images. *IEEE Trans. Med. Imaging* 39, 2000–2012. doi:10.1109/TMI.2019.2962792

Toikkanen, M., Kwon, D., and Lee, M. (2021). "ReSGAN: Intracranial hemorrhage segmentation with residuals of synthetic brain CT scans," in 2021 Medical Image Computing Computer Assisted Intervention, 400–409.

Vijayan, M., and Reddy, P. H. (2016). Peripheral biomarkers of stroke: Focus on circulatory microRNAs. *Biochim. Biophys. Acta* 1862, 1984–1993. doi:10.1016/j.bbadis.2016.08.003

Wang, J. D., Sun, K., Cheng, T. H., Jiang, B. R., Deng, C. R., Zhao, Y., et al. (2021). Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3349–3364. doi:10.1109/TPAMI.2020.2983686

Wang, S., Liang, K. M., Pan, C. W., Ye, C. Y., Li, X. L., Liu, P., et al. (2020). "Segmentation-based method combined with dynamic programming for brain midline delineation," IEEE International Symposium on Biomedical Imaging, 772–776.

Wu, H., Wang, W., Zhong, J., Lei, B., Wen, Z., and Qin, J. (2021). SCS-net: A scale and context sensitive network for retinal vessel segmentation. *Med. Image Anal.* 70, 102025. doi:10.1016/j.media.2021.102025

Xu, J., Zhang, R., Zhou, Z., Wu, C., Gong, Q., Zhang, H., et al. (2021). Deep network for the automatic segmentation and quantification of intracranial hemorrhage on CT. *Front. Neurosci.* 14. doi:10.3389/fnins.2020.541817

Yu, F., Koltun, V., and Funkhouser, T. (2017). "Dilated residual networks," IEEE/CVF Conference on Computer Vision and Pattern Recognition, 636–644.

Yu, N. N., Yu, H., Li, H. N., Ma, N. N., Hu, C. N., and Wang, J. (2022). A robust deep learning segmentation method for hematoma volumetric detection in intracerebral hemorrhage. *Stroke* 53, 167–176. doi:10.1161/STROKEAHA.120.032243

Zhao, W. B., Wu, C. J., Stone, C., Ding, Y. C., and Ji, X. M. (2020). Treatment of intracerebral hemorrhage: Current approaches and future directions. *J. Neurol. Sci.* 416, 117020. doi:10.1016/j.jns.2020.117020

Zhu, Y., Pan, X., Zhu, J., and Li, L. (2020). Multi-scale strategy based 3D dual-encoder brain tumor segmentation network with attention mechanism, in. 2020 IEEE International Conference on Bioinformatics and Biomedicine, 952–957.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 25 November 2022
DOI 10.3389/fgene.2022.982019

Check for updates

OPEN ACCESS

EDITED BY
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China

REVIEWED BY
Xiannian Zhang,
Capital Medical University, China
Vahid Shahrezaei,
Imperial College London, United Kingdom

*CORRESPONDENCE
Aritro Nath,
anath@coh.org

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 30 June 2022
ACCEPTED 04 November 2022
PUBLISHED 25 November 2022

CITATION
Bishara I, Chen J, Griffiths JI, Bild AH and Nath A (2022), A machine learning framework for scRNA-seq UMI threshold optimization and accurate classification of cell types.
Front. Genet. 13:982019.
doi: 10.3389/fgene.2022.982019

COPYRIGHT
© 2022 Bishara, Chen, Griffiths, Bild and Nath. 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 machine learning framework for scRNA-seq UMI threshold optimization and accurate classification of cell types

Isaac Bishara[1,2], Jinfeng Chen[1,3], Jason I. Griffiths[1], Andrea H. Bild[1] and Aritro Nath[1]*

[1]Department of Medical Oncology and Therapeutics, City of Hope Comprehensive Cancer Center, Duarte, CA, United States, [2]Irell & Manella Graduate School of Biological Science, City of Hope Comprehensive Cancer Center, Duarte, CA, United States, [3]State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

Recent advances in single cell RNA sequencing (scRNA-seq) technologies have been invaluable in the study of the diversity of cancer cells and the tumor microenvironment. While scRNA-seq platforms allow processing of a high number of cells, uneven read quality and technical artifacts hinder the ability to identify and classify biologically relevant cells into correct subtypes. This obstructs the analysis of cancer and normal cell diversity, while rare and low expression cell populations may be lost by setting arbitrary high cutoffs for UMIs when filtering out low quality cells. To address these issues, we have developed a novel machine-learning framework that: 1. Trains cell lineage and subtype classifier using a gold standard dataset validated using marker genes 2. Systematically assess the lowest UMI threshold that can be used in a given dataset to accurately classify cells 3. Assign accurate cell lineage and subtype labels to the lower read depth cells recovered by setting the optimal threshold. We demonstrate the application of this framework in a well-curated scRNA-seq dataset of breast cancer patients and two external datasets. We show that the minimum UMI threshold for the breast cancer dataset could be lowered from the original 1500 to 450, thereby increasing the total number of recovered cells by 49%, while achieving a classification accuracy of >0.9. Our framework provides a roadmap for future scRNA-seq studies to determine optimal UMI threshold and accurately classify cells for downstream analyses.

KEYWORDS

ScRNA-seq, UMI (unique molecular identifier), QC, quality control, threshold, optimization, gene, cut off

Introduction

One of the key objectives in cancer genomics is characterizing the composition and diversity of cancer and normal cells in the tumor microenvironment (TME) (Ren et al., 2018). Several studies have shown that the composition of the TME, such as the prevalence of infiltrating lymphocytes, polarity of myeloid cells and signaling from stromal components play a critical role in the maintenance and progression of malignant cells, and can serve as indicators of therapeutic potential and response (Gooden et al., 2011; Awad et al., 2018; Maibach et al., 2020; Wu et al., 2020; Geng et al., 2021). The study of the TME has been greatly enhanced by the introduction of single cell RNA sequencing (scRNA-seq), which enabled characterizing the diversity and phenotypes of cells in a tumor at a fine resolution (Rubio-Perez et al., 2021; Tang et al., 2022).

Since the introduction of scRNA-seq more than a decade ago, several incremental technological advances have improved the accessibility and quality of transcriptomic analyses (Hwang et al., 2018; Chen et al., 2019). One such advance is the introduction of unique molecular identifiers (UMIs) which allows direct quantification of available transcripts (Islam et al., 2013). While non-UMI scRNA-seq platforms as Smart-Seq2 provide an improved transcript coverage and high level of mappable reads, UMI platforms such as 10X and drop-seq benefit from the limited amplification bias from highly abundant transcripts (Picelli et al., 2014; Zhang et al., 2019). The higher throughput of UMI platforms also improves the detection rates of rare cell populations, such as certain immune cells, within tumor samples (Azizi et al., 2018). Thus, scRNA-seq technologies have greatly enhanced the ability to characterize the diversity of cancer cells and the TME.

However, the ability to accurately classify the cell types in scRNA-seq dataset is often limited by technical factors, such as read quality of the cells. The quality control (QC) process in a typical scRNA-seq pipeline involves identification and filtering out cells of low quality, typically based on the number of UMIs, number of unique genes, and/or the percentage of mitochondrial DNA (mtDNA). The stress induced by droplet-based UMI methods introduces a challenge in ensuring that the UMIs map to healthy cells (Chittur et al., 1988). For example, cells with leaky or damaged membranes can result in a drop in the number of UMIs and genes detected, while the number of UMIs mapping mtDNA may become relatively high (Luecken and Theis, 2019). This complicates the distinction between true low-quality cells and quiescent, small, and/or rare cell populations, thus creating a trade-off between cell quality and diversity during the QC process (Luecken and Theis, 2019).

Since mitochondrial DNA content varies significantly across organisms and tissues, comprehensive analysis of these variables helps to establish universal organism and tissue-specific threshold guidelines (Osorio and Cai, 2021). However, due to the variability in the number of UMI and genes owing to biological and technical factors, a similar universal threshold cannot be established *a priori*. A probabilistic model was proposed to sort out low-quality cells but its accuracy was limited by the prevalence of low-quality cells, which is usually unknown (Hippen et al., 2021). Additionally, several scRNA-seq pre-processing pipelines included different approaches for QC including the option to view the UMI distribution per cell type using user-defined marker genes (McCarthy et al., 2017; Guo et al., 2021; Grandi et al., 2022). However, these approaches generally depend on the user's judgment to detect outliers (low-quality cells) from reads and/or gene distribution curve. The scRNA-seq literature shows the number of reads threshold selected at QC can vary from as low as 100 and up to 2500 UMIs, yet the rationale for selecting such thresholds is usually missing (Liu et al., 2021; Gambardella et al., 2022; Gao et al., 2022; Karademir et al., 2022; Lian et al., 2022). Another approach which involves an iterative process between the QC step and downstream analysis was also proposed to improve the detection of low-quality cells (Luecken and Theis, 2019). But the mechanism by which the downstream information can be used to optimize an initial reads threshold is not yet defined.

To address the lack of a systematic approach to determine an optimal reads threshold for filtering cells and classifying cells with high accuracy, we have developed a novel machine learning framework that uses cell identity information collected from a high-quality gold standard. Using this approach, we can identify the lowest reads cut-off that can be implemented in an scRNA-seq data and accurately classify cell lineages and subtypes. We used expert-labelled lineage and cell type identities from a gold standard breast cancer scRNA-seq dataset to train the predictive classifiers. We systematically downsampled the reads per cell in the gold standard dataset using a Poisson model and then applied the classifier to predict cell types. We then calculated the prediction accuracies of the classifiers using the known identities of the cells. This allowed us to determine the optimal threshold at which sufficient biological information was retained. Using this approach, we rescued 49% more cells from the gold standard dataset, which is valuable for downstream analyses of the TME. Using two external datasets, we show that our approach can be applied to low expression cells and to subtypes of major cell types as neutrophils and T-cell subtypes, respectively. Importantly, our framework can be extended to any scRNA-seq dataset where users seek to rescue and classify additional cells at optimal read depths.

Methods

Analysis workflow

The analysis pipeline consists of the following main steps (Figure 1). We applied a stringent QC threshold on the FELINE dataset (raw UMIs) to filter for the high-confidence, high-quality

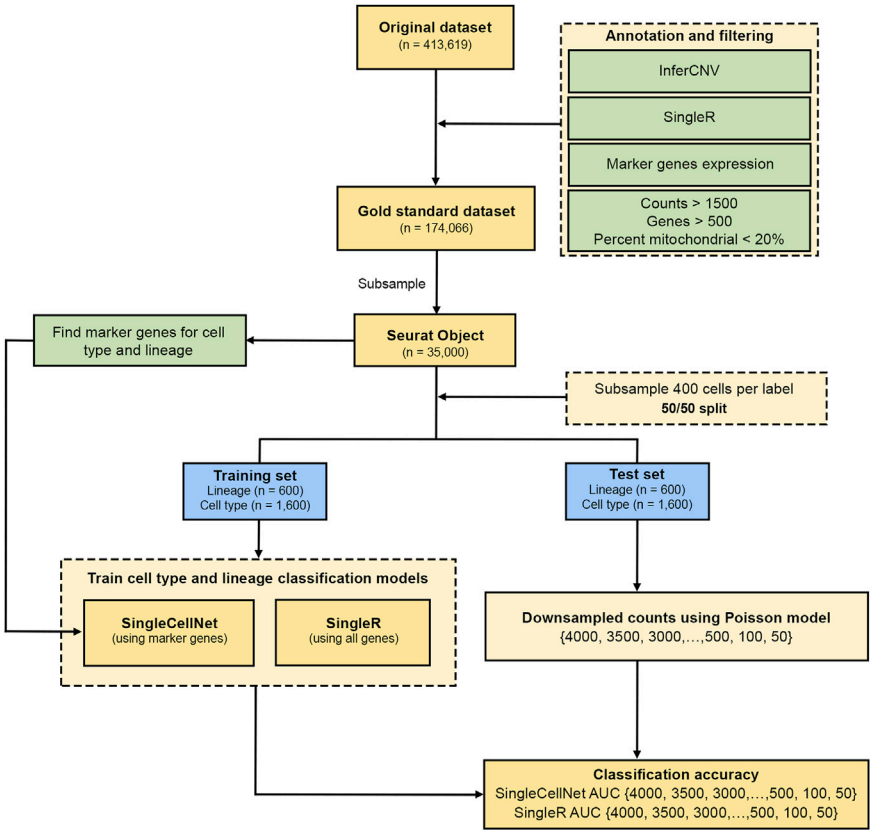


FIGURE 1
Analysis plan workflow. Flow chart shows the process of initial QC and generation of gold standard cell type annotations from the FELINE dataset. This is followed by a 50/50 split of a subsample into training and test sets for both SingleR and SingleCellNet classifiers for all datasets. The test set counts were then transformed using a Poisson model using different thresholds which is then used to determine the classification accuracy of lineage and cell type labels.

cells. A combination of unsupervised and supervised expert-led approaches was used to generate the high-quality cell lineage and subtype labels which were used at the gold standard for downstream analysis. For each dataset, we first split it into training and test sets (50/50). Next, the training set was used to train the classification models to predict cell lineage and subtypes. The test set was then downsampled using Poisson model at different target UMI thresholds. We then assessed the accuracy of the classification models on the test set at different target UMI thresholds. The analysis steps are described in more details in the subsections below.

Gold standard scRNA-seq dataset pre-processing

We used the FELINE clinical trial scRNA-seq dataset which spans 35 patients with ER-positive HER2-negative early stage breast cancer (Griffiths et al., 2021). The patient samples were processed using the 10X Chromium platform and sequenced using 150-bp paired-end sequencing at a median depth of 34,000 reads per cell (Griffiths et al., 2021). The reads were aligned to a reference genome (GRChg38) using Bioinformatics the ExperT SYstem and CellRanger v.3.0.2 pipelines (Chen and Chang, 2017). FeatureCounts was then used to generate a matrix of gene transcript UMIs for each cell, which we refer to as "original dataset" in this manuscript (Liao et al., 2014).

To generate the gold standard dataset, we applied a stringent QC filter which retained cells with >1,500 reads, 500—7,000 unique genes, and less than 20% mitochondrial content, as reported in the original study (Griffiths et al., 2021). After filtering out "low-quality" cells and doublets, we retained 176,644 "high-quality" cells. To generate Uniform Manifold Approximation and Projection (UMAP), we log-normalized, scaled the count matrix, and ran principal component analysis (PCA) on the 2000 highly variable genes using R package Seurat v.4.1.1 (Butler et al., 2018). We then

constructed the K nearest neighbor and using Seurat's FindNeighbor function on 10 principal components which was used to construct the UMAP. We then used SingleR to generate a preliminary cell type label for each cell using Human Primary Cell Atlas (HPCA) as a reference (Mabbott et al., 2013; Aran et al., 2019). These labels were used to annotate the clusters as either epithelial, stromal, or immune based on the most frequent cell type labels by SingleR. The SingleR labels were validated using lineage marker gene expression for epithelial cells (*KRT19, CDH1*), stromal cells (*FAP, HTRA1*), and immune cells (*PTPRC*) (Griffiths et al., 2021). SingleR cell type labels were also validated using cell type marker gene expression for macrophages (*CSF1R, CD163*), T-cells (*CD2, CD247*), B-cells (*MS4A1, IGHM*), fibroblasts (*COL5A1, FBLN1*), endothelial cells (*VWF*), pericytes (*RGS5*), and adipocytes (*CIDEA*). To identify putative cancer cell, we used InferCNV which predicts copy number alterations based on the positional gene expression intensity across all chromosomes (Korsunsky et al., 2019). We used stromal and immune cells as normal references for InferCNV and labelled epithelial cells with positive copy number alterations (CNA) profile as cancer cells (Griffiths et al., 2021). All downstream analyses excluded non-malignant epithelial cells. The raw (un-normalized) UMI count matrix of the gold standard dataset was used for model training and assessment. A random unbiased subsample of the gold standard dataset (n = 35,000) was used to create a Seurat object for downstream analysis. We removed cells with >15,000 reads to account for any missed doublets.

External datasets

In addition to the FELINE dataset, we used a subset of whole blood scRNA-seq dataset (GSE163668) which we will refer to as "Combes dataset" (Combes et al., 2021). We combined 3 pooled libraries (GSM4995425, GSM4995426, GSM4995427) spanning 8 patients, removed RBCs and used the remaining cells with the authors' cell type labels in our analysis. We also used a PeripheralBlood Mononuclear Cells (PBMC) dataset freely available from 10X Genomics which we will refer to as the "PBMC dataset" (10x Genomics, 2016). We processed this dataset as described in "Seurat-Guided Clustering Tutorial" (Hoffman et al., 2022). Cells with more than 5% mitochondrial counts or more than 2,500 genes or less than 200 genes were filtered out. After clustering the cells, cell types were annotated using the canonical markers as follows: Naive $CD4^+$ T (*IL7R, CCR7*), $CD14^+$ Mono (*CD14, LYZ*), Memory $CD4^+$ (*IL7R, S100A4*), B cells (*MS4A1*), $CD8^+$ T (*CD8A*), FCGR3A + Mono (*FCGR3A, MS4A7*), NK (*GNLY, NKG7*), DC (*FCER1A, CST3*), Platelet (PPBP).

Low-quality cells subset

For "low-quality" cells which that were excluded from the gold standard dataset, we predicted the cell type labels using SingleR and human primary cell atlas (HPCA) as a reference (Mabbott et al., 2013; Aran et al., 2019). To generate lineage labels, we aggregated cell type predictions into lineage labels as follows: epithelial (epithelial cells), stromal (fibroblasts, endothelial cells, chondrocytes, osteoblast, smooth muscles), immune (T-cells, B-cells, macrophages, monocytes, NK cells, neutrophils). To study the outcome of the initial and optimized thresholds on cell retention rate, we combined the gold standard subsample (n = 35,000) with a low-quality subsample (n = 35,000) for a total of 70,000 cells.

Training lineage and cell subtype classification modes

We used two different multi-class prediction algorithms for the analysis, SingleCellNet (SCN) and SingleR. SCN is a Random Forest classifier developed for scRNA-seq datasets and implemented as R package singleCellNet v.0.1.0 (Tan and Cahan, 2019). SingleR is a reference-based cell type classifier where after an internal marker genes identification step, cell identity is determined by Spearman correlation between the expression profile of the unknown cell and the reference samples e.g., HPCA (Aran et al., 2019). Due to the infeasibility to train a random forest classifier on all genes, we applied Seurat's FindAllMarkers function (test.use = "negbinom", min.pct = 0.5, max.cells.per.ident = 2000, logfc.threshold = 0.5) to generate lineage and cell type marker gene sets. For either lineage or cell type levels, we sampled 400 cells per label using splitCommon function implemented in R package singleCellNet v.0.1.0. The lineage and cell type samples were split 1:1 into a training and test set. For the SCN classifier, the UMI matrices of both training sets were filtered for the corresponding marker gene set previously identified. The SCN classifier was trained using scn_train function (nTopGenes = 100, nRand = 50, nTrees = 1000, nTopGenePairs = 200) implemented in the singleCellNet package. In contrast, the SingleR classifier was trained on all available genes in UMI matrices without filtering using trainSingleR function implemented in the R package SingleR v.1.6.1.

Systematic downsampling of reads and genes

To simulate reduce average reads per cell at a pre-specified threshold, we downsampled the reads from high-quality cells. We used a Poisson distribution model to calculate a transformation factor. The probabilities density function for an integer vector x is defined as:

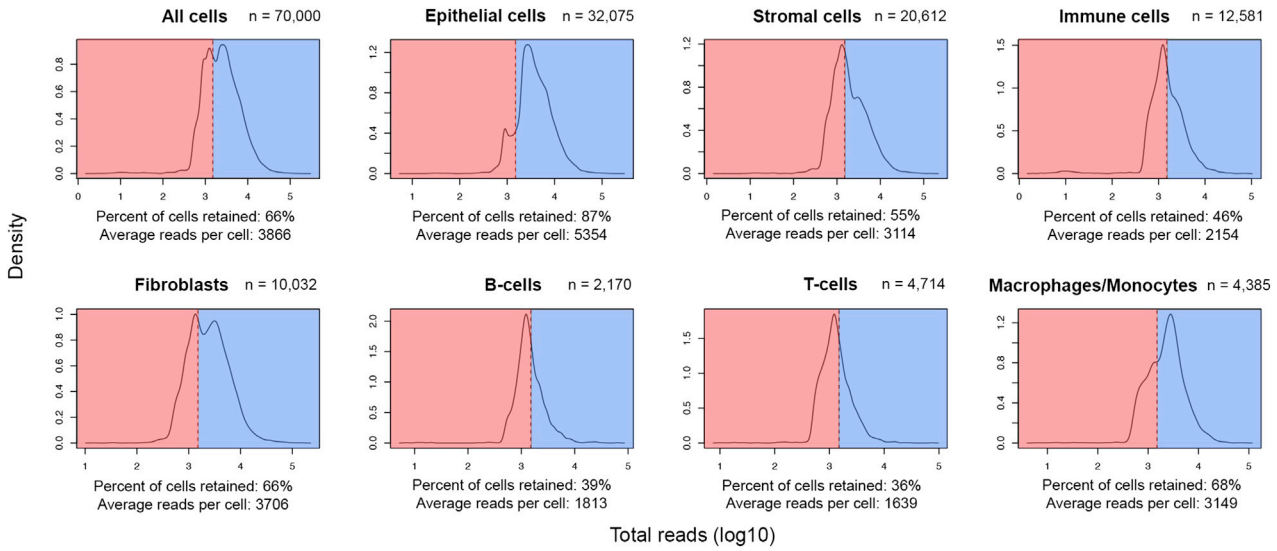


FIGURE 2
Post-QC retention rate varies across different lineages and cell types in the FELINE dataset. Density plots depict the reads-per-cell distribution across different lineages and cell types within a subsample of the original dataset (n = 70,000). The initial QC count cut-off (1,500 reads), as dashed line, splits the fraction of cells considered as "high-quality", highlighted in blue, from the cells considered as "low-quality", highlighted in red, across different cell populations. The average count and the fraction of "high-quality" cells are annotated for each population.

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where, λ is the point mass (Poisson rate). For each cell, we generated a vector of random deviates of length = number of genes, and λ = target threshold/total reads. Reads from each cell were multiplied by their transformation factor to reduce the total counts per cell to the desired threshold.

To downsample the genes of the FELINE dataset, we first converted the UMI matrix into binary expression. For cells where n> = 1, we reduced random n genes from being expressed to not expressed (1 → 0) where n is the number of genes above test threshold. Each transformed matrix was then used to assess the accuracy of classification for the corresponding threshold. In the non-binary experiments, the remaining binary matrix was converted back to a non-binary UMI matrix for assessment while in binary-experiments, both the training and downsampled matrices were binary.

Model assessment

Using the SCN and SingleR trained models, we generate the predicted labels for all downsampled matrices using scn_predict and classifySingleR functions, respectively. We then used the true labels to calculate the Area Under Receiver Operating Characteristic Curve (AUROCC) for both models at each threshold using the R package pROC v.1.18.0.

Results

Cell retention rates in gold standard scRNA-seq dataset

The diversity of cell populations within the TME introduces a challenge when applying a UMI threshold across tumor samples: a stringent, high UMI threshold would remove most of the low-quality cells, but also lose important populations with low reads like immune cells. In contrast, a lenient threshold would retain the low-UMI populations, but this could also increase the noise and possibly skewing the results of the downstream analysis. In addition, the QC step is usually performed early in the analysis pipeline where biological information (cell identities) is not yet available. Thus, a biology-driven revision of QC thresholds can be easily overseen. In the FELINE dataset, we had used 1,500 reads as a threshold for low-quality cells (Figure 1) (Griffiths et al., 2021). To construct the gold standard dataset, we used InferCNV to identify cancer cells and SingleR to predict normal cell identities which were verified by marker gene expression (Supplementary Figures S1A,B).

After meticulous cell type labelling of high-quality cells, a closer view of UMI distribution across cell lineages showed a high level of retention of epithelial cells (87%) post-QC. In contrast, only around half of the stromal and immune cells were retained (Figure 2). As breast cancer cells are of epithelial origin (Noureen et al., 2022), it is expected that actively

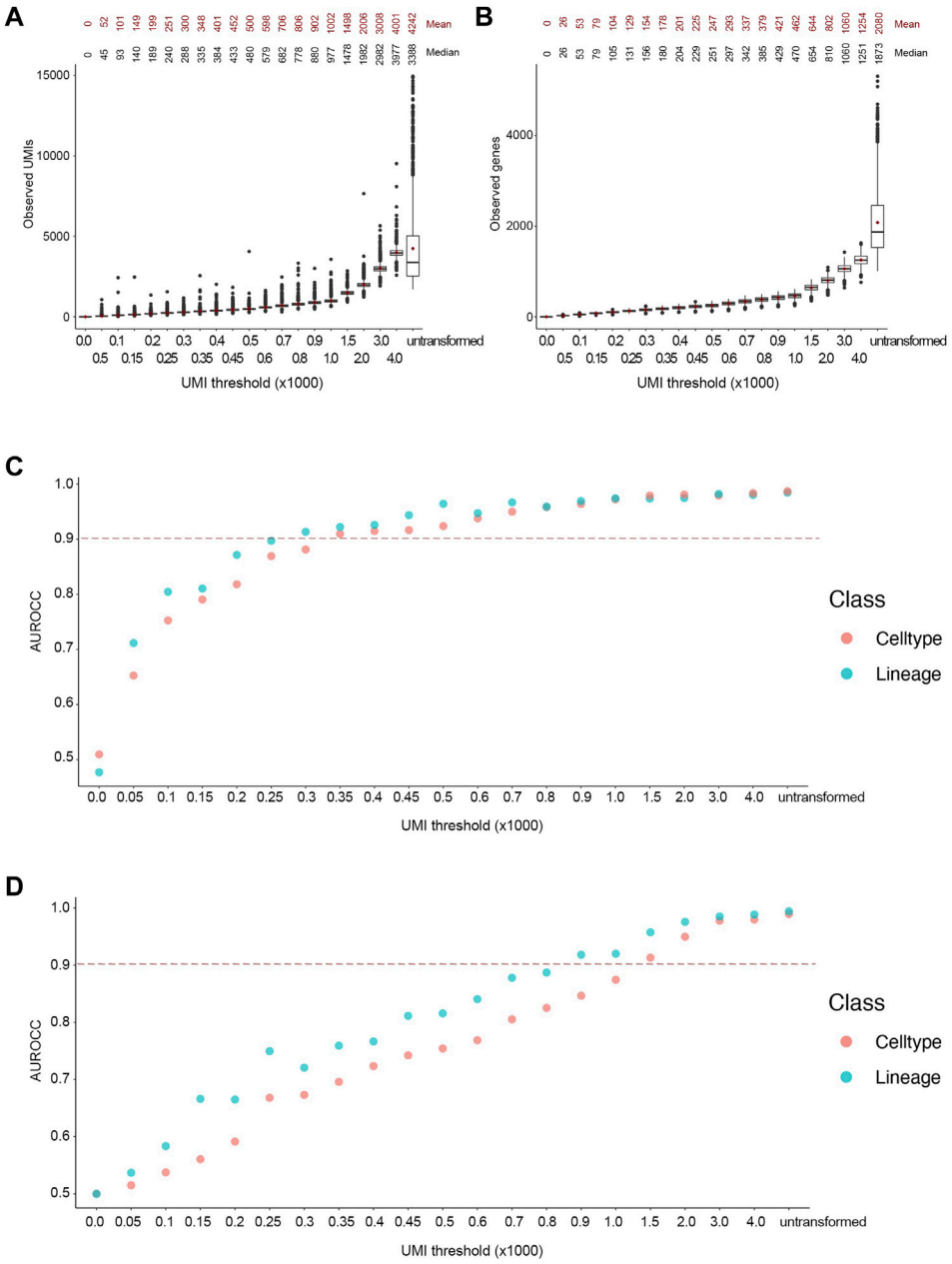


FIGURE 3
Accurate lineage and cell type classification at 450 UMIs in the FELINE dataset. **(A,B)** Boxplot showing the post-transformation distributions of observed UMIs **(A)** and number of unique genes **(B)** across all thresholds and untransformed control. Mean and median values for each distribution are denoted. **(C,D)** Area under the receiver operating characteristic curve (AUROCC) values are shown for the raw (untransformed) counts as well as the downsampled counts at different thresholds using the SingleR model **(C)** and the SingleCellNet model **(D)**. The AUROCC values for both lineage and cell type assessments are shown for each model as well as the selected AUROCC cut-off value (>0.9), dashed line.

proliferating cancer cells were driving a higher average UMI among epithelial cells (5,354 UMIs) than stromal (3,114 UMIs) or immune cells (2,154 UMIs) (Figure 2). In addition, at the finer cell subtype annotation level, two-thirds of macrophages/monocytes were retained, while only a third of the sequenced population of T and B lymphocytes were retained (Figure 2). Since B- and T-lymphocytes have the lowest average UMIs per cell in this cohort (1,813 and 1,639 respectively), the initial QC threshold only retained a small fraction of these cells for downstream analyses, suggesting an optimization of the initial threshold might be required.

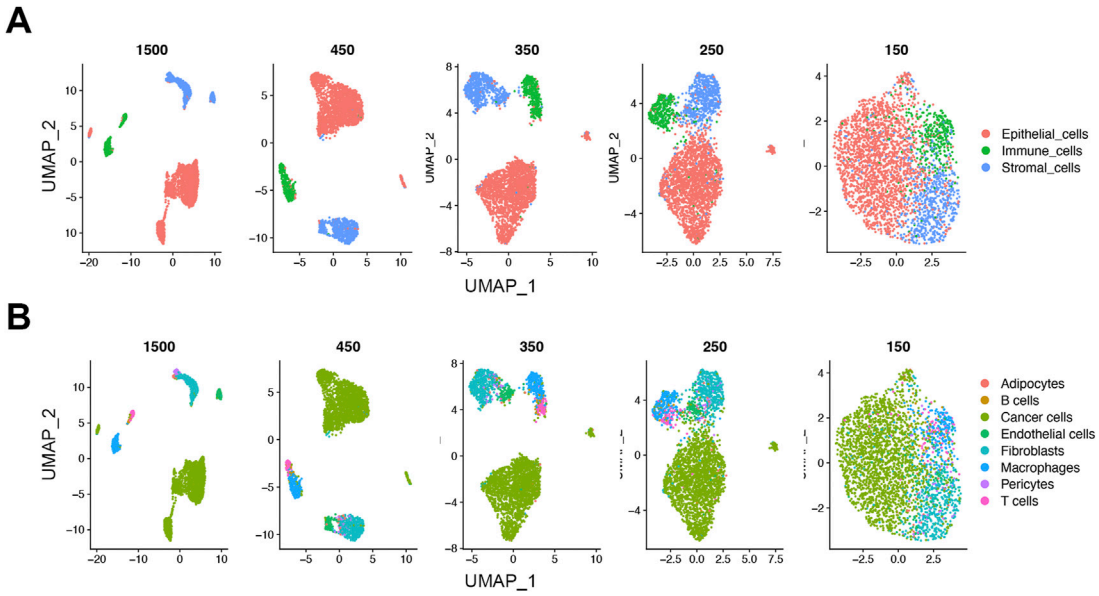


FIGURE 4
Loss of distinct cell clusters on UMAP below 450 UMIs in the FELINE dataset. Dimension reduction using Uniform Manifold Approximation and Projection (UMAP) shows that as count thresholds fall below 450 reads, a gradual loss of the distinct cell clusters is observed on lineage **(A)**, and cell type levels **(B)** (n = 1,500).

Machine learning framework guides threshold optimization and accurate classification

We developed a novel framework that systematically identified the lowest read depth threshold that can be used to accurately classify cell lineages and subtypes. Our approach trained classifiers for lineage and subtypes on a training subset of the gold standard dataset, and then predicted the cell lineage and subtypes of a held-out test or validation subset from the gold standard dataset at progressively diminished read depths. By following this approach, we could identify what is the minimum number of average reads required to accurately classify cells.

We used SCN and SingleR multi-class prediction algorithms to determine the lowest UMI threshold where sufficient biological signal was retained. We then applied a Poisson model to the test datasets to downsample to a set of desired reads threshold including 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000 and 4000 UMIs.

Following the transformation, the mean number of UMIs in the downsampled cells were close to the desired UMI thresholds (Figure 3A). Indeed, the reads in the downsampled cells followed a Poisson distribution, as the variance increased at higher thresholds. Noticeably, the number of unique genes followed a Poisson distribution as well (Figure 3B). We used the trained classifiers to predict lineage and cell type labels for the downsampled cells. The ground truth and predicted labels were used to generate a confusion matrix to calculate the area under the receiver operator curve (AUROCC) at each threshold. We considered AUROCC values above 0.9 to be accurate classifications. The SingleR classifier showed an accurate prediction of both lineage and cell types at an average read depth of 450 UMIs or ~200 genes (Figure 3C). However, the model progressively lost its predictive ability at below the 250 UMIs threshold. On the other hand, the SCN classifier showed an accurate prediction for both classes at an average read depth of 1,500 UMIs or ~650 genes, while its predictive ability was gradually lost at thresholds below 800 UMIs (Figure 3D). The accuracy of the SingleR classifier relatively plateaued at the 350 UMI threshold. However, the accuracy of the SCN classifier increased linearly throughout with the increasing thresholds. As expected, almost all the AUROCC values for the broader lineage class were equal or higher than the narrower cell type class. It's worth mentioning that SingleR classifier showed an overall higher classification accuracy which we attribute to the fact that SingleR calculates the spearman correlation between each cell's expression profile and reference cells regardless of expression values while SCN only considers expressed genes e.g., non-zero expression values. Consequently, we selected the conservative 450 UMIs from the more accurate classifier at the finer cell type resolution as the optimized threshold.

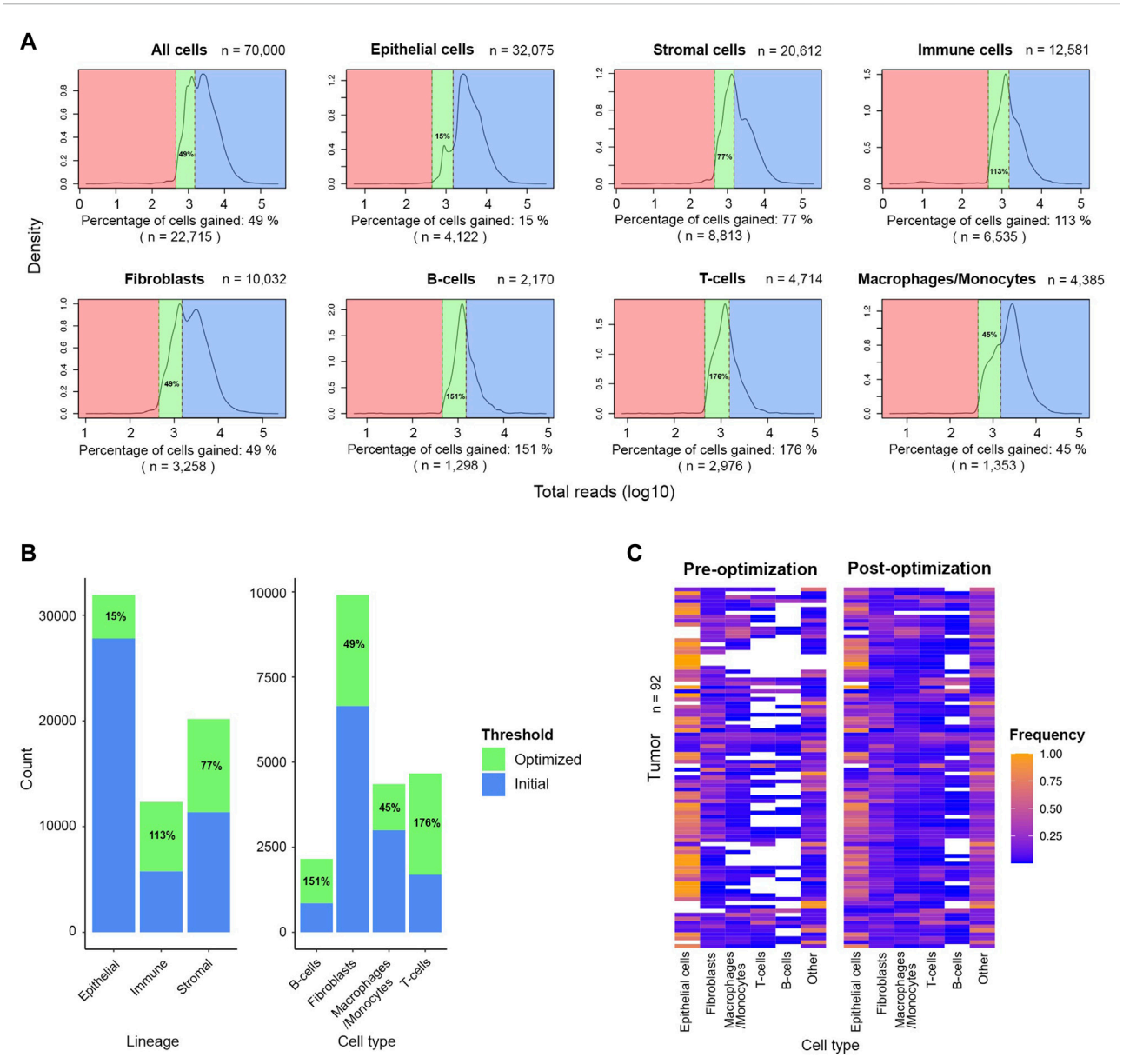


FIGURE 5
Significant number of stromal and immune cells are rescued after applying the optimized threshold of 450 UMIs in the FELINE dataset. **(A)** Density plots shows the UMI distribution across lineages and cell types within high- and low-quality cells subset (*n* = 70,000). The initial threshold (1,500 UMIs), dashed line to the right, and the optimized threshold (450 UMIs), dashed line to the left, are shown for each plot. The initial "high-quality" cells, the rescued cells after applying the revised cut-off, and the low-quality cells are highlighted in blue, green, and red, respectively. The fraction and number of cells gained relative to initially retained cells is denoted under each plot. **(B)** Bar plot showing the cell number and percentage gain for lineage and cell types after applying the optimized threshold. **(C)** Heatmap showing the relative frequency of different cell types before and after applying the optimized UMI threshold of 450 in 92 tumor samples.

In addition, we performed downsampling of gene numbers by dropping random genes at different maximum number of genes thresholds (Supplementary Figures S2A,B). Like the UMI downsampling, accurate classification (AUROCC >0.9) of lineages and cell types was achieved using 200 and 600 genes for SingleR and SCN classifier, respectively (Supplementary Figures 2C,D). We then applied the same transformation to a binary count matrix for training and test sets (Supplementary Figures S3A,B). Both classifiers yielded similar performance to non-binary counts at 250 and 450 genes for SingleR and SCN, respectively (Supplementary Figures S3C,D). Given the typical correlation between observed between UMIs and number of genes, it was not surprising that similar thresholds were obtained using the UMI-based and the gene number approaches.

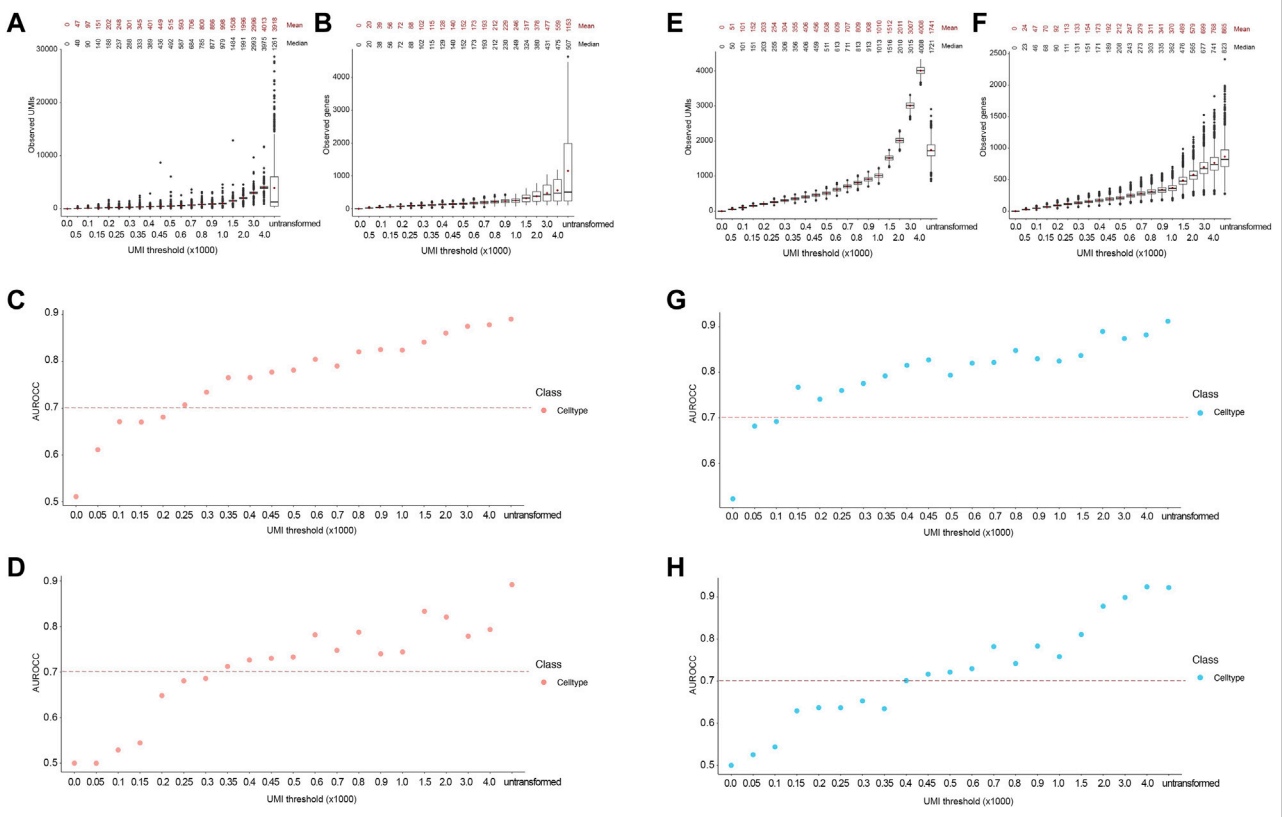


FIGURE 6
Accurate lineage and cell type classification at 250 and 150 UMIs in the Combes and PBMC datasets, respectively. For Combes dataset, **(A,B)** Boxplot showing the post-transformation distributions of observed UMIs **(A)** and number of unique genes **(B)** across all thresholds and untransformed control. Mean and median values for each distribution are denoted. **(C,D)** Area under the receiver operating characteristic curve (AUROCC) values are shown for the raw (untransformed) counts as well as the downsampled counts at different thresholds using the SingleR model **(C)** and the SingleCellNet model **(D)**. For PBMC dataset, **(E, F)** Boxplot showing the post-transformation distributions of observed UMIs **(E)** and number of unique genes **(F)** across all thresholds and untransformed control. Mean and median values for each distribution are denoted. **(G,H)** Area under the receiver operating characteristic curve (AUROCC) values are shown for the raw (untransformed) counts as well as the downsampled counts at different thresholds using the SingleR model **(G)** and the SingleCellNet model **(H)**. The AUROCC values for cell type assessment are shown for each model as well as the selected AUROCC cut-off value (>0.7), dashed line.

Loss of distinct clustering below the optimized threshold

To see the effect of downsampling on the low dimensional data structure, we analyzed the downsampled cells from the 1500, 450, 350, 250, and 150 read thresholds using uniform manifold approximation and projections (UMAPs). Similar to the initial 1500 UMI threshold, the cells at the 450 UMI threshold showed distinct separate clusters at the lineage level (Figure 4A). As threshold was reduced, the inter-cluster distances gradually decreased. On the cell type level, the cells at the 450-threshold not only clustered by lineage but retained a rational biological hierarchy as shown by subtype cluster grouping (Figure 4B). As with the lineage level, the distinct clustering was gradually lost at lower thresholds (Figures 4A,B). This suggests that biological information retained at as low as 450 reads-per-cell maintains cell identity in our dataset.

Optimized QC threshold rescue substantial number of cells with low transcription level

To increase the number of stromal and immune cells available for downstream analysis, we applied the optimized threshold of 450 reads-per-cell to a subsample of the original dataset (n = 70,000). Relative to number of cells retained by the initial threshold of 1,500 reads, the optimized threshold rescued an additional 8,813 stromal cells and 6,535 immune cells, an increase of 77% and 113%, respectively (Figures 5A,B). The gain was even more prominent among the cells with low average reads as

2,976 T-cells and 1,298 B-cells were rescued which is 176% and 151%, respectively, more cells compared to the populations retained by the initial threshold. The gain among fibroblasts and macrophages/monocytes was also notable as the initial populations increased by more than 40% after applying the optimized threshold. The inclusion of rescued cells markedly improved the representation of diversity across all tumor samples, previously dominated by epithelial cells (Figure 5C). With the new thresholds, we observed a notable gain in lymphocytes across several tumors. We also noted that the optimized threshold led to the gain of 10 additional tumor samples that were excluded by the initial threshold. Thus, threshold optimization allowed the re-evaluation of cells initially penalized and discarded for their natively low expression. These rescued cells can then be incorporated in downstream analysis to characterize the TME.

Applications in datasets containing cells with low expression and fine-grain labels

To test the applicability of our approach to cell types with low gene expression, we used the Combes dataset (see Methods), which contains cell types with low expression levels, including as neutrophils and platelets. As with the FELINE dataset, we applied the transformation based on Poisson distribution to systematically downsample the counts in the Combes dataset. The resultant UMI means were reflective of the desired target UMI thresholds (Figures 6A,B). Using the original published cell type labels as ground truth, the cell type classification AUROCC for the untransformed counts were about 0.9, reflecting the low average read depth of this dataset (1599 UMIs) and very low coverage in some cell types, such as neutrophils (621 UMIs) and platelets (740 UMIs). SingleR achieved AUROCC >0.7 for this dataset at 250 UMIs or ~90 genes while SCN achieved this level of accuracy at 350 UMIs or ~115 genes (Figures 6C,D).

Similarly, we used the 10X PBMC dataset test (see methods for details) to demonstrate that the application of the framework in cell types with fine-grain labels. The PBMC dataset (average 2371 UMIs) contains fine-grain classification of monocytes and T cells. In addition to $CD14^+$ and FCGR3A + monocytes, this dataset contains different T cells subtypes like naïve $CD4^+$, memory $CD4^+$, and $CD8^+$ T cells. Again, we applied the transformation based on Poisson distribution to systematically downsample and obtain resultant UMIs that were reflective of the desired target thresholds (Figures 6E,F). SingleR classified cells with AUROCC >0.7 at 150 UMIs or ~70 genes threshold, while the SCN classifier achieved this level of accuracy at 400 UMIs or ~170 genes (Figures 6G,H). Taken together, these results demonstrate that our framework can be applied to datasets containing cell types with low expression and fine granularity.

Discussion

Single cell RNA-seq of tumor samples have proved indispensable for TME studies. This has allowed researchers to perform analyses such as in-depth classification of the composition of tumors, identifying the key signaling mechanisms operating in cancer and non-cancer cells and characterizing the heterogeneity and evolution of cancer cells, which were not previously feasible using bulk-RNA sequencing (Nath and Bild, 2021). However, the detection of rare cell populations among the diverse TME is limited by the number of cells the scRNA-seq platform can handle. The introduction of UMI-based platforms allowed for higher cell capacity which better captures the diversity of the TME. However, arbitrary UMI thresholding during the standard scRNA-seq QC risks losing considerable number of cells, such as immune cells with low expression. This can lead to inaccurate assessment of the composition of the TME and overlook critical associations between diversity and tumor traits. For example, the presence of cytotoxic T cells in the TME is strongly associated immunotherapy response in multiple cancers (Sade-Feldman et al., 2018; Kim et al., 2021; Nagasaki et al., 2022). Therefore, assessment of immune response based on diversity of infiltrating lymphocytes could improve by optimizing the UMI thresholds. Recent studies to characterize the communication networks between various individual cell types within breast tumor have revealed unique signaling networks operate in tumors resistant or sensitive to cell cycle inhibitor therapy (Griffiths et al., 2022). Resolving these communication links also requires optimizing the UMI thresholds to ensure that the TME measured using scRNA-seq reflects the true composition of the tumor.

To develop a framework that enables optimization of UMI thresholds, we used a systematic approach to downsample UMIs and accurately classify cells by lineage and cell type. We trained two classifiers, SCN and SingleR, on expert-labelled subsample of our gold standard FELINE dataset which was originally filtered using a stringent UMI threshold. We then downsampled the FELINE dataset using a Poisson transformation and evaluated the classification accuracies at various thresholds. Using a conservative AUROCC >0.9 as the cut-off for accurate classification in the FELINE dataset, we determined a significantly lower new threshold at 450 UMIs, corresponding to slightly more than 200 genes, compared to the initial threshold at 1,500 UMIs. The optimized threshold retrieved substantial number of additional cells that were initially disposed-off during filtering. The gain was prominent among cells with lower average reads than cancer cells such as stromal and immune cells. Notably, B- and T-lymphocytes populations increased more than 150% by applying the optimized threshold. We also noticed that the downsampled cells at this threshold retained similar distinct clustering patterns across lineages and cell type groups on the UMAP as the gold standard dataset. However, this was not the case at lower thresholds where the inter-cluster

distances were gradually lost. We also explored gene downsampling using random gene removal at different thresholds using binary and non-binary input which resulted in similar optimal threshold to the UMI downsampling.

We further extend the application of our framework to two additional datasets. Analyses with the Combes dataset revealed that cells with low average expression, like neutrophils, can also be used in our framework to optimize thresholds. Similarly, analyses with the PBMC dataset showed that fine grain classification of cells can be accommodated in the framework.

While this approach improved the diversity of major lineages and cell types of the FELINE, Combes and PBMC datasets, its current application depends on the original labeling accuracy for cell identities. This can be challenging for some cell populations, such as cells that lack established RNA markers. Currently, the framework relies on reliable labeling of cell types in the high-quality cells. A future addition to this framework could integrate additional biological information such as pathway level information and molecular signatures to identify biologically relevant clusters and improve classification accuracy.

Our machine learning framework provides a systematic approach to optimize the initial UMI/reads threshold commonly used in scRNA-seq pipelines based on cell type annotations of cells with high read depth. This is especially valuable in rescuing cells with natively low expression like immune cells. Optimizing the QC reads threshold significantly improves the efficiency of cell diversity TME studies while maintaining accurate classification of lineage and cell type. Notably, this framework can be applied to any scRNA-seq dataset where rescuing rare or low expression cells is crucial for downstream analysis.

Data availability statement

The Combes et al. data are available through Gene Expression Omnibus under accession code GSE163668. The PBMC data are available at https://www.10xgenomics.com/resources/datasets. Other datasets and code used in this analysis are available on our GitHub repository at https://github.com/ibishara/scRNA-seq_threshold_optimization.

Author contributions

IB: Conceptualization; Data curation; Software; Formal analysis; Validation; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. JC: Formal analysis. JG: Formal analysis. AB: Resources; Supervision; Funding acquisition; Project administration; Writing—review and editing. AN: Conceptualization; Supervision; Investigation; Methodology; Project administration; Writing—original draft; Writing—review and editing.

Funding

Research reported in this publication was supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award number U54CA209978 and U01CA264620 awarded to AHB, and a pilot grant awarded to AN under U54CA209978. Work performed in the Integrative Genomics Core at City of Hope was supported by the NCI of the NIH under award number P30CA33572.

Acknowledgments

The authors would like to acknowledge the work provided by the Leadership and Staff of the CoH Center for Informatics most notably Research Informatics, and the utilization of the POSEIDON data exploration, visualization, and analysis platform including the Honest Broker process.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.982019/full#supplementary-material

References

10x Genomics (2016). pbmc3k-Datasets. Available at: https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k?.

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20 (2), 163–172. doi:10.1038/s41590-018-0276-y

Awad, R. M., De Vlaeminck, Y., Maebe, J., Goyvaerts, C., and Breckpot, K. (2018). Turn back the TIMe: Targeting tumor infiltrating myeloid cells to revert cancer progression. *Front. Immunol.* 9, 1977. doi:10.3389/fimmu.2018.01977

Azizi, E., Carr, A. J., Plitas, G., Cornish, A. E., Konopacki, C., Prabhakaran, S., et al. (2018). Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* 174 (5), 1293–1308. doi:10.1016/J.CELL.2018.05.060

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36 (5), 411–420. doi:10.1038/NBT.4096

Chen, G., Ning, B., and Shi, T. (2019). Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* 10, 317. doi:10.3389/fgene.2019.00317

Chen, X., and Chang, J. T. (2017). Planning bioinformatics workflows using an expert system. *Bioinformatics* 33 (8), 1210–1215. doi:10.1093/BIOINFORMATICS/BTW817

Chittur, K. K., Mcintire, L. V., and Rich, R. R. (1988). Shear stress effects on human T cell function. *Biotechnol. Prog.* 4 (2), 89–96. doi:10.1002/BTPR.5420040207

Combes, A. J., Courau, T., Kuhn, N. F., Hu, K. H., Ray, A., Chen, W. S., et al. (2021). Global absence and targeting of protective immune states in severe COVID-19. *Nature* 591 (7848), 124–130. doi:10.1038/S41586-021-03234-7

Gambardella, G., Viscido, G., Tumaini, B., Isacchi, A., Bosotti, R., and di Bernardo, D. (2022). A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nat. Commun.* 13 (1), 1–12. doi:10.1038/s41467-022-29358-6

Gao, S., Wu, Z., Arnold, B., Diamond, C., Batchu, S., Giudice, V., et al. (2022). Single-cell RNA sequencing coupled to TCR profiling of large granular lymphocyte leukemia T cells. *Nat. Commun.* 13 (1), 1–14. doi:10.1038/s41467-022-29175-x

Geng, X., Chen, H., Zhao, L., Hu, J., Yang, W., Li, G., et al. (2021). Cancer-associated fibroblast (CAF) heterogeneity and targeting therapy of CAFs in pancreatic cancer. *Front. Cell Dev. Biol.* 9, 1766. doi:10.3389/fcell.2021.655152

Gooden, M. J. M., de Bock, G. H., LeffersN.Daemen, T., and Nijman, H. W. (2011). The prognostic influence of tumour-infiltrating lymphocytes in cancer: A systematic review with meta-analysis. *Br. J. Cancer* 105 (11), 93–103. doi:10.1038/bjc.2011.189

Grandi, F., Caroli, J., Romano, O., Marchionni, M., Forcato, M., and Bicciato, S. (2022). popsicleR: A R package for pre-processing and quality control analysis of single cell RNA-seq data. *J. Mol. Biol.* 434 (11), 167560. doi:10.1016/J.JMB.2022.167560

Griffiths, J. I., Chen, J., Cosgrove, P. A., O'Dea, A., Sharma, P., Ma, C., et al. (2021). Serial single-cell genomics reveals convergent subclonal evolution of resistance as early-stage breast cancer patients progress on endocrine plus CDK4/6 therapy. *Nat. Cancer* 2 (626), 658–671. doi:10.1038/s43018-021-00215-7

Griffiths, J. I. (2022). *Cancer cells communicate with macrophages to prevent T cell activation during development of cell cycle therapy resistance*, 2022. bioRxiv. doi:10.1101/2022.09.14.507931

Guo, W., Wang, D., Wang, S., Shan, Y., Liu, C., and Gu, J. (2021). scCancer: a package for automated processing of single-cell RNA-seq data in cancer. *Brief. Bioinform.* 22 (3), bbaa127. doi:10.1093/BIB/BBAA127

Hippen, A. A., Falco, M. M., Weber, L. M., Erkan, E. P., Zhang, K., Doherty, J. A., et al. (2021). miQC: An adaptive probabilistic framework for quality control of single-cell RNA-sequencing data. *PLoS Comput. Biol.* 17 (8), e1009290. doi:10.1371/JOURNAL.PCBI.1009290

Hoffman, P. (2022) Seurat - guided clustering tutorial ● Seurat. Available at: https://satijalab.org/seurat/articles/pbmc3k_tutorial.html (Accessed: 31 October 2022).

Hwang, B., Lee, J. H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50 (88), 96–14. doi:10.1038/s12276-018-0071-8

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., et al. (2013). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11 (22), 163–166. doi:10.1038/nmeth.2772

Karademir, D., Todorova, V., Ebner, L. J. A., Samardzija, M., and Grimm, C. (2022). Single-cell RNA sequencing of the retina in a model of retinitis pigmentosa reveals early responses to degeneration in rods and cones. *BMC Biol.* 20 (11), 86–19. doi:10.1186/S12915-022-01280-9

Kim, M., Min, Y. K., Jang, J., Park, H., Lee, S., and Lee, C. H. (2021). Single-cell RNA sequencing reveals distinct cellular factors for response to immunotherapy targeting CD73 and PD-1 in colorectal cancer. *J. Immunother. Cancer* 9 (7), e002503. doi:10.1136/JITC-2021-002503

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16 (12), 1289–1296. doi:10.1038/S41592-019-0619-0

Lian, Q., Zhang, K., Zhang, Z., Duan, F., Guo, L., Luo, W., et al. (2022). Differential effects of macrophage subtypes on SARS-CoV-2 infection in a human pluripotent stem cell-derived model. *Nat. Commun.* 13 (11), 2028–2114. doi:10.1038/s41467-022-29731-5

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* 30 (7), 923–930. doi:10.1093/BIOINFORMATICS/BTT656

Liu, Y., He, S., Wang, X. L., Peng, W., Chen, Q. Y., Chi, D. M., et al. (2021). Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat. Commun.* 12 (11), 741–818. doi:10.1038/s41467-021-21043-4

Luecken, M. D., and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* 15 (6), e8746. doi:10.15252/MSB.20188746

Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C., and Hume, D. A. (2013). An expression atlas of human primary cells: Inference of gene function from coexpression networks. *BMC Genomics* 14 (1), 1–13. doi:10.1186/1471-2164-14-632

Maibach, F., Sadozai, H., Seyed Jafari, S. M., Hunger, R. E., and Schenk, M. (2020). Tumor-infiltrating lymphocytes and their prognostic value in cutaneous melanoma. *Front. Immunol.* 11, 2105. doi:10.3389/fimmu.2020.02105

McCarthy, D. J., Campbell, K. R., Lun, A. T. L., and Wills, Q. F. (2017). Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33 (8), 1179–1186. doi:10.1093/BIOINFORMATICS/BTW777

Nagasaki, J., Inozume, T., Sax, N., Ariyasu, R., Ishikawa, M., Yamashita, K., et al. (2022). PD-1 blockade therapy promotes infiltration of tumor-attacking exhausted T cell clonotypes. *Cell Rep.* 38 (5), 110331. doi:10.1016/J.CELREP.2022.110331

Nath, A., and Bild, A. H. (2021). Leveraging single-cell approaches in cancer precision medicine. *Trends Cancer* 7 (4), 359–372. doi:10.1016/J.TRECAN.2021.01.007

Noureen, N., Ye, Z., Chen, Y., Wang, X., and Zheng, S. (2022). Signature-scoring methods developed for bulk samples are not adequate for cancer single-cell RNA sequencing data. *eLife* 11, e71994. doi:10.7554/ELIFE.71994

Osorio, D., and Cai, J. J. (2021). Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics* 37 (7), 963–967. doi:10.1093/BIOINFORMATICS/BTAA751

Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9 (11), 171–181. doi:10.1038/nprot.2014.006

Ren, X., Kang, B., and Zhang, Z. (2018). Understanding tumor ecosystems by single-cell sequencing: Promises and limitations. *Genome Biol.* 19 (11), 211–214. doi:10.1186/S13059-018-1593-Z

Rubio-Perez, C., Planas-Rigol, E., Trincado, J. L., Bonfill-Teixidor, E., Arias, A., Marchese, D., et al. (2021). Immune cell profiling of the cerebrospinal fluid enables the characterization of the brain metastasis microenvironment. *Nat. Commun.* 12 (11), 1503–1510. doi:10.1038/s41467-021-21789-x

Sade-Feldman, M., Yizhak, K., Bjorgaard, S. L., Ray, J. P., de Boer, C. G., Jenkins, R. W., et al. (2018). Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell* 175 (4), 998–1013.e20. doi:10.1016/J.CELL.2018.10.038

Tan, Y., and Cahan, P. (2019). SingleCellNet: A computational tool to classify single cell RNA-seq data across platforms and across species. *Cell Syst.* 9 (2), 207–213. doi:10.1016/J.CELS.2019.06.004

Tang, P. C. T., Chung, J. Y. F., Xue, V. W. W., Xiao, J., Meng, X. M., Huang, X. R., et al. (2022). Smad3 promotes cancer-associated fibroblasts generation via macrophage–myofibroblast transition. *Adv. Sci.* 9 (1), 2101235. doi:10.1002/ADVS.202101235

Wu, K., Lin, K., Li, X., Yuan, X., Xu, P., Ni, P., et al. (2020). Redefining tumor-associated macrophage subpopulations and functions in the tumor microenvironment. *Front. Immunol.* 11, 1731. doi:10.3389/fimmu.2020.01731

Zhang, X., Li, T., Liu, F., Chen, Y., Yao, J., Li, Z., et al. (2019). Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Mol. Cell* 73 (1), 130–142. doi:10.1016/J.MOLCEL.2018.10.020

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fmolb.2022.937242

Check for updates

OPEN ACCESS

EDITED BY
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-Sen University, China

REVIEWED BY
Ho Seok Seo,
Catholic University of Korea, South Korea
Cristiana Tanase,
Victor Babes National Institute of Pathology (INCDVB), Romania
Brandon Zagorski,
BMZ Analytics, Canada

*CORRESPONDENCE
Chong Feng,
ryanfeng@live.com
Tai Ma,
matai@ahmu.edu.cn

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

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 06 May 2022
ACCEPTED 15 November 2022
PUBLISHED 01 December 2022

CITATION
Zhang C, Zhang Y, Yang Y-H, Xu H, Zhang X-P, Wu Z-J, Xie M-M, Feng Y, Feng C and Ma T (2022), Machine learning models for predicting one-year survival in patients with metastatic gastric cancer who experienced upfront radical gastrectomy.
Front. Mol. Biosci. 9:937242.
doi: 10.3389/fmolb.2022.937242

COPYRIGHT
© 2022 Zhang, Zhang, Yang, Xu, Zhang, Wu, Xie, Feng, Feng and Ma. 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.

Machine learning models for predicting one-year survival in patients with metastatic gastric cancer who experienced upfront radical gastrectomy

Cheng Zhang[1,2†], Yi Zhang[1†], Ya-Hui Yang[1†], Hui Xu[1,2], Xiao-Peng Zhang[3], Zhi-Jun Wu[4], Min-Min Xie[1], Ying Feng[1], Chong Feng[3]* and Tai Ma[1]*

[1]Department of Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China, [2]Anhui Provincial Cancer Institute/Anhui Provincial Office for Cancer Prevention and Control, Hefei, Anhui, China, [3]Department of Noncommunicable Diseases and Health Education, Hefei Center for Disease Control and Prevention, Hefei, Anhui, China, [4]Department of Oncology, Ma'anshan Municipal People's Hospital, Ma'anshan, Anhui, China

Tumor metastasis is a common event in patients with gastric cancer (GC) who previously underwent curative gastrectomy. It is meaningful to employ high-volume clinical data for predicting the survival of metastatic GC patients. We aim to establish an improved machine learning (ML) classifier for predicting if a patient with metastatic GC would die within 12 months. Eligible patients were enrolled from a Chinese GC cohort, and the complete detailed information from medical records was extracted to generate a high-dimensional dataset. Appropriate feature engineering and feature filter were conducted before modeling with eight algorithms. A 10-fold cross validation (CV) nested in a holdout CV (8:2) was employed for hyperparameter tuning and model evaluation. Model selection was based on the area under the receiver operating characteristic (AUROC) curve, recall, and precision. The selected model was globally explained using interpretable surrogate models. Of the total 399 cases (median survival of 8.2 months), 242 patients survived less than 12 months. The linear discriminant analysis (LDA), support vector machine (SVM), and random forest (RF) model had the highest AUROC (0.78 $\pm$ 0.021), recall (0.93 $\pm$ 0.031), and precision (0.80 $\pm$ 0.026), respectively. The LDA model created a new function that generally separated the two classes. The predicted probability of the SVM model was interpreted using a linear regression model visualized by a nomogram. The predicted class of the RF model was explained using a decision tree model. In summary, analyzing high-volume medical data by ML is helpful to produce an improved model for predicting the survival in patients with metastatic GC. The algorithm should be carefully selected in different practical scenarios.

KEYWORDS

stomach neoplasms, neoplasm metastasis, survival analysis, supervised machine learning, electronic medical record, clinical laboratory information system

Introduction

Gastrectomy with adequate lymphadenectomy provides a potential opportunity of cure for resectable gastric cancer (GC) (Smyth et al., 2020); however, a substantial proportion of patients still develop recurrence or metastasis afterward (Chen et al., 2021; Hisamori et al., 2021). The prognosis of metastatic GC is expected to be poor; the survival time after relapse varies from 3–15 months (Smyth et al., 2020), depending on the metastatic site (Chau et al., 2004; Lee et al., 2007; Kim et al., 2008; Custodio et al., 2017), performance status (Chau et al., 2004; Lee et al., 2007; Kim et al., 2008; Custodio et al., 2017), palliative chemotherapeutic regimen, and other factors (Custodio et al., 2017; Zhu et al., 2022). Several models have been established based on clinical trials or real-world data, aiming to precisely estimate the survival probability in these patients (Chau et al., 2004; Lee et al., 2007; Kim et al., 2008; Koo et al., 2011; Custodio et al., 2017). Although different sets of variables have been incorporated, the ability of survival prediction in the traditional model is dissatisfactory. A Spanish multicenter study (the AGAMENON study) developed a nomogram-based model to predict the survival of patients with advanced GC (AGC), with an accuracy of 0.67 in the validation set (Custodio et al., 2017). A Korean single-center study constructed a score-based model with an accuracy of 0.58 (Koo et al., 2011), then externally validated another three models (Chau et al., 2004; Lee et al., 2007; Kim et al., 2008), and showed similar performances in the same population (Koo et al., 2011). Our previous work also developed a score-based model in a Chinese cohort with a c-index of 0.67 (Ma et al., 2021). Meanwhile, we validated seven published models (Chau et al., 2004; Lee et al., 2007; Kim et al., 2008; Koo et al., 2011; Takahari et al., 2014; Wang et al., 2016; Kim et al., 2020) in a Chinese population, and the results showed that the area under receiver operating characteristic (AUROC) curves was only about 0.60 (Xu et al., 2021).

The traditional prognostic model is frequently built by the logistic or Cox regression analysis on the basis of the well-known clinical and pathological variables, for example, performance status (Chau et al., 2004; Lee et al., 2007; Kim et al., 2008; Takahari et al., 2014), tumor differentiation (Custodio et al., 2017; Kim et al., 2020), metastatic sites (Kim et al., 2008; Takahari et al., 2014; Wang et al., 2016), and routine laboratory tests (Koo et al., 2011; Custodio et al., 2017; Kim et al., 2020; Ma et al., 2021). The selection of candidate variables is typically guided by the clinical experience and previous literature. In the era of digital medicine, the electronic medical record (EMR) and laboratory information system (LIS) make massive medical data readily available; nevertheless, we are still far from taking full advantage of them. One possible reason is the incompetence of the classic statistical method in dealing with numerous independent variables, which emphasizes the need for adopting a new strategy of statistics.

Machine learning (ML) is increasingly used for data mining due to its capacity to tackle big data. In order to utilize the abundant digital medical records and further improve model performance in predicting the survival of GC patients with recurrence or metastasis after radical gastrectomy, we enroll eligible participants from a retrospective GC cohort, build a high-dimensional dataset from the EMR and LIS, identify the most relevant prognostic factors, and implement modeling using several ML algorithms.

Materials and methods

Study setting and population

In this retrospective study, we trained ML models using different algorithms to predict if a GC patient would die within 12 months after the first metastasis or recurrence because 12 months is typically recognized as the median survival time for patients with AGC (Smyth et al., 2020). The participants were enrolled from a registered hospital-based GC cohort (ChiCTR1800019978, http://www.chictr.org.cn/). The consecutive gastric or esophagogastric junction carcinoma patients who underwent radical gastrectomy and developed disease recurrence or metastasis were included and followed up in the cohort. Those patients with multiple primary malignant tumors or with no records of laboratory examinations at the time of metastasis were excluded. The EMR and LIS were retrieved to obtain data for analysis. The survival information was acquired from the death register system or by telephonic follow-up conducted every 3 months. The overall survival (OS) was defined as the interval between the first metastasis and death or the last follow-up. The workflow of the study is illustrated in Figure 1.

All procedures performed in the study involving human participants were in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The studies involving human participants were reviewed and approved by the Ethics Committee of The First Affiliated Hospital of Anhui Medical University (reference number: Quick-PJ-2021-05-19). The Ethics Committee waived the requirement of written informed consent for participation.

Dataset and feature engineering

All features are listed in Supplementary Table S1. Briefly, the dataset included information about demography, histopathology, surgical resection, postoperative adjuvant chemotherapy, first-line palliative chemotherapy, radiotherapy, baseline laboratory records at the time of metastasis (routine blood test,

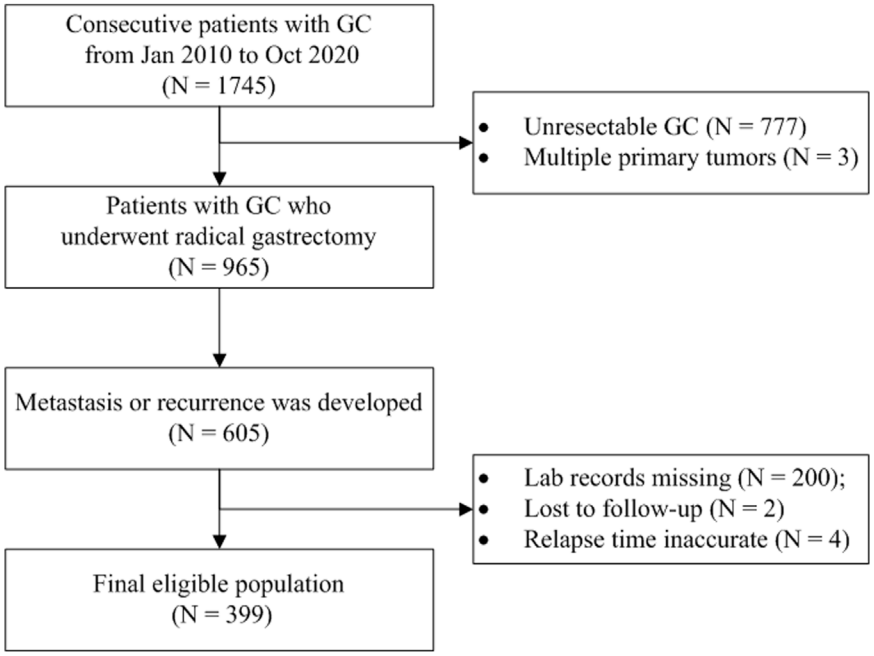


FIGURE 1
Flow diagram of patient selection. Abbreviation: GC, gastric cancer.

biochemistry, coagulation, immunology, and tumor biomarkers), and survival. Each aspect had several items to record the details, so a high-dimensional dataset was generated.

Categorical features were transformed by one-hot encoding. Numerical features were standardized, normally transformed, or grouped where appropriate. In our dataset, missing values generally occurred at random, so they were deleted (the fraction of the missing values over the total cases was more than 30%) or imputed using decision tree algorithm.

ML model performance may suffer from high dimensionality, so, here, some features were filtered out prior to modeling. A feature with zero or near-zero variance was first dropped because it provided no useful information to a model. The rule of detecting a near-zero variance feature was (Smyth et al., 2020) that the fraction of unique values over the sample size was less than 10% and (Hisamori et al., 2021) the ratio of the frequency of the most prevalent value to the frequency of the second most prevalent value was more than 20% (Boehmke and Greenwell, 2019). Next, we used the importance value calculated by the random forest (RF) algorithm to rank the features and select a number of them that contributed most to the model. The specific number was tuned by a random search during model development. All these data-dependent preprocessings were conducted in isolation of each resampling iteration in order to avoid data leakage.

Model development

First, the entire dataset was randomly split into a training set and a validation set (8:2) as the outer layer. Then, the training set was further randomly split by 10-fold cross-validation (CV) as the inner layer. The inner layer was used to tune hyperparameters by random search, and the best configuration was passed on the validation set from the outer layer to evaluate the model performance. The nested CV design reduced the risk of overestimation of the model because the information of the training set was not leaked into the validation set. The whole process was repeated five times for averaging the effect of randomness, so we used the mean value to measure the model performance. The AUROC curve was the primary indicator to evaluate the model because it did not have any bias toward classifiers on balanced or imbalanced binary prediction problems (He and Ma, 2013). Precision and recall were also crucial as they reflected the false-positive error and the false-negative error of the model, respectively. In addition, accuracy and F1 score (the harmonic mean of precision and recall) were also calculated.

We used eight common classification algorithms for modeling: kernel K-nearest neighbor (KKNN), linear discriminant analysis (LDA), support vector machine (SVM), RF, XGBoost, ridge regression, LASSO regression, and elastic net regression. For each algorithm, the hyperparameters that needed to be tuned and the optimal settings are given in Supplementary

Table S2. The whole project was deployed using RStudio 1.4.1717 with packages "mlr3verse" (modeling pipeline and framework), "kknn" (KKNN algorithm), "e1071" (SVM algorithm), "MASS" (LDA algorithm), "ranger" (RF algorithm), "xgboost" (XGBoost algorithm), and "glmnet" (ridge, LASSO, and elastic net regression). To make a comparison with the traditional method, we used the logistic regression as a reference algorithm.

A learning curve is used to diagnose if the sample size is adequate for modeling and if an overfitting or underfitting problem occurs. It comprises two lines that represent the errors of the training set and the validation set, respectively, in relation to the sample size. The training learning curve shows how well the model is learning, and the validation curve shows how well the model is generalizing. If a model is underfitting, the error of the training set is close to that of the validation set, so obtaining more samples is unlikely to improve the performance. In contrast, if a model is overfitting, the gap between the errors of the training set and validation set is large, so adding more samples is likely to be helpful.

Model interpretation

Only the selected models were interpreted, which comprised the LDA, the SVM, and the RF-based model. The model interpretation was based on the final model built on the entire dataset with the tuned hyperparameters or model configuration. The general theories of the three algorithms were briefly demonstrated. The LDA aims to learn a new line, called the discriminant function (DF) that combines the original features in a linear fashion, weighting greater for "better" predictors and less for "poorer" predictors. The value that gives the weight for each feature is called the DF coefficient, which indicates how much it contributes to class discrimination. The DF separates the centroid of each class (OS longer than or equal to 12 months *versus* shorter than 12 months in the case) by maximizing the difference between the class centroids and minimizing the within-class variance when the data being projected onto the DF (Rhys, 2020).

The SVM and the RF algorithms are more alike "black-box" models. The SVM algorithm finds an optimal linear hyperplane that best separates the two classes and is penalized for having cases inside its decision boundary defined by the support vectors. The algorithm can also add a kernel, namely, an extra dimension, to deform the feature space, so that a linear hyperplane can separate the classes (Rhys, 2020). The RF algorithm is an implementation of a bagging technique for decision tree algorithm. It randomly samples cases and features to create a large number of tree classifiers on a binary prediction task that are highly uncorrelated. Then, new data are passed to the trees to make their own prediction, and the model prediction is made based on the majority of the predictions from each tree (Rhys, 2020).

Global surrogate is a common global model-agnostic method to interpret a black box model (e.g., SVM or RF model) by using a surrogate model with a good intuition. In this case, we train a linear regression model or a decision tree model to fit the black box-predicted probability or response, respectively. R-square was used to measure how close the surrogate model is to the black box model (Elshawi et al., 2019).

Results

As shown in Figure 1, 399 GC patients developing metastasis or recurrence after curative intent gastrectomy were enrolled for modeling. The median survival after metastasis was 8.2 months. Two lost to follow-up cases were removed. Fourteen patients were still alive (survival time ranged from 20.7 to 144.0 months, median 95.1 months), so all the living patients had an OS of no less than 12 months and were assigned to the negative subgroup. Of 385 patients who reached the endpoint (mOS = 7.8 months), 143 patients survived for no less than 12 months (negative subgroup). Overall, the negative subgroup consisted of 157 cases, and the positive subgroup (post-metastatic survival time <12 months) consisted of 242 cases. The ratio of the majority to the minority was 1.54:1. The five most frequent metastasis sites were 46.1% for distant lymph nodes (n = 184), 26.8% for the liver (n = 107), 19.8% for the peritoneum (n = 79), 15.8% for bone (n = 63), and 15.5% for the chest (n = 62). The baseline information is briefly presented in Table 1. No missing value existed.

After excluding some features with massive missingness, a total of 62 laboratory indexes were in the feature space, and the distribution of missing values was generally at random (Supplementary Figure S1A). No clear pattern was observed when comparing these indexes between the positive and the negative subgroups (Supplementary Figure S1B,C). The feature filter process returned a feature list ordered by the RF importance and then passed it into further modeling. Interestingly, the highly important features were generally the laboratory information at metastasis, i.e., inflammatory index, blood cell test, and biochemistry. Supplementary Figure S1C illustrates the 30 most important features for brevity.

After tuning the optimal number of the features and the hyperparameters, the model performances are shown in Table 2. The LDA model had the highest AUROC (m ± sd; 0.78 ± 0.021), followed by the SVM model with an AUROC of 0.77 ± 0.014 and the RF model with an AUROC of 0.77 ± 0.0064. The SVM model ranked first with respect to recall (0.93 ± 0.031), and the RF model ranked first with respect to precision (0.80 ± 0.026). So, we further look into these three models.

The LDA itself had no hyperparameter to be tuned, so only the number of the included features should be tested. Figure 2A shows that the inclusion of the first 25 or 40 features yielded the lowest error in the validation set, so 25 was chosen to make the

TABLE 1 Baseline characteristics of the patients enrolled for modeling.

	Overall (n = 399)	Negative (n = 157)	Positive (n = 242)	*p*-value
General information				
Sex (male)	283 (70.93)	108 (68.79)	175 (72.31)	0.519
Age at surgery, y	62.00 (54.00, 68.50)	60.00 (51.00, 68.00)	64.00 (56.00, 69.00)	0.012
DFS, mo	11.87 (5.94, 21.72)	11.43 (4.40, 21.70)	12.02 (6.56, 21.70)	0.378
Age at first metastasis, y	64.00 (55.00, 70.00)	62.00 (53.00, 69.00)	65.00 (56.00, 71.00)	0.014
OS, mo	8.23 (3.77, 17.20)	21.40 (14.90, 31.13)	4.46 (2.51, 7.34)	<0.001
Pathological information				
T stage				0.048
T1	12 (3.01)	6 (3.82)	6 (2.48)	
T2	26 (6.52)	17 (10.83)	9 (3.72)	
T3	231 (57.89)	89 (56.69)	142 (58.68)	
T4a	105 (26.32)	38 (24.20)	67 (27.69)	
T4b	21 (5.26)	7 (4.46)	14 (5.79)	
Tx	4 (1.00)	0 (0.00)	4 (1.65)	
N stage				<0.001
N0	59 (14.79)	38 (24.20)	21 (8.68)	
N1	81 (20.30)	41 (26.11)	40 (16.53)	
N2	105 (26.32)	28 (17.83)	77 (31.82)	
N3a	112 (28.07)	36 (22.93)	76 (31.40)	
N3b	39 (9.77)	14 (8.92)	25 (10.33)	
Nx	3 (0.75)	0 (0.00)	3 (1.24)	
Grade				0.093
G1	5 (1.25)	2 (1.27)	3 (1.24)	
G2	84 (21.05)	43 (27.39)	41 (16.94)	
G3	273 (68.42)	98 (62.42)	175 (72.31)	
G4	5 (1.25)	3 (1.91)	2 (0.83)	
Gx	32 (8.02)	11 (7.01)	21 (8.68)	
Location[¶]				
Cardia	223 (55.89)	93 (59.24)	130 (53.72)	0.327
Body	137 (34.34)	51 (32.48)	86 (35.54)	0.603
Pylorus	125 (31.33)	44 (28.03)	81 (33.47)	0.301
Linitis plastica	4 (1.00)	1 (0.64)	3 (1.24)	>0.999
Histology[¶]				
Adenocarcinoma, NOS	333 (83.46)	136 (86.62)	197 (81.40)	0.218
Mucinous adenocarcinoma	54 (13.53)	14 (8.92)	40 (16.53)	0.043
SRC	25 (6.27)	7 (4.46)	18 (7.44)	0.323
Borrmann type				0.819
I	18 (4.51)	8 (5.10)	10 (4.13)	
II	111 (27.82)	42 (26.75)	69 (28.51)	
III	223 (55.89)	93 (59.24)	130 (53.72)	
IV	25 (6.27)	9 (5.73)	16 (6.61)	
Unknown	22 (5.51)	5 (3.18)	17 (7.02)	
Treatment information				
Resection site				0.731
Proximal	16 (4.01)	8 (5.10)	8 (3.31)	
Distal	101 (25.31)	35 (22.29)	66 (27.27)	
Total	273 (68.42)	112 (71.34)	161 (66.53)	
Others	9 (2.26)	2 (1.27)	7 (2.89)	

(Continued on following page)

TABLE 1 (*Continued*) Baseline characteristics of the patients enrolled for modeling.

	Overall (n = 399)	Negative (n = 157)	Positive (n = 242)	*p*-value
Procedure				0.557
Open gastrectomy	364 (91.23)	141 (89.81)	223 (92.15)	
Laparoscopic gastrectomy	34 (8.52)	16 (10.19)	18 (7.44)	
Unknown	1 (0.25)	0 (0.00)	1 (0.41)	
Lymphadenectomy				0.695
D1	128 (32.08)	54 (34.39)	74 (30.58)	
D2	212 (53.13)	83 (52.87)	129 (53.31)	
Unknown	59 (14.79)	20 (12.74)	39 (16.12)	
Radiotherapy (yes)	4 (1.00)	1 (0.64)	3 (1.24)	0.939
Adjuvant chemotherapy (yes)	298 (74.69)	116 (73.89)	182 (75.21)	0.858
Adjuvant chemotherapy cycles[§]	5 (3, 6)	5 (3, 6)	5 (3, 6)	0.290
Palliative chemotherapy (yes)	317 (79.45)	137 (87.26)	180 (74.38)	0.003
First-line drugs[¶]				
Platinum	45 (11.28)	26 (16.56)	19 (7.85)	0.012
Fluorouracil	72 (18.05)	33 (21.02)	39 (16.12)	0.267
Taxane	28 (7.02)	14 (8.92)	14 (5.79)	0.319

Abbreviations: DFS, disease-free survival; SRC, signet-ring cell.
[¶] In this variable, the count of each category did not sum up to the total number of cases within a subgroup due to overlapping distribution; therefore, a chi-squared or Fisher's exact test was conducted within each row. Otherwise, the test for categorical variables was conducted within each matrix.
[§] Only the patients who had a history of adjuvant chemotherapy were summarized.
Continuous variables were presented by the median (interquartile range).

TABLE 2 Comparison of model performance across different machine learning algorithms for predicting 12-month survival in patients with metastatic gastric cancer.

Algorithm	AUROC	Recall	Precision	F1-score	Accuracy
LR	0.68 ± 0.055	0.76 ± 0.058	0.69 ± 0.083	0.72 ± 0.033	0.68 ± 0.045
ER	0.75 ± 0.006	0.74 ± 0.031	0.76 ± 0.037	0.75 ± 0.033	0.70 ± 0.041
KNN	0.75 ± 0.014	0.85 ± 0.039	0.74 ± 0.014	0.79 ± 0.017	0.73 ± 0.018
LASSO	0.76 ± 0.016	0.74 ± 0.027	0.76 ± 0.046	0.75 ± 0.035	0.70 ± 0.046
LDA	**0.78 ± 0.021**	0.84 ± 0.018	0.78 ± 0.023	0.81 ± 0.016	0.76 ± 0.023
RF	0.77 ± 0.006	0.82 ± 0.039	**0.80 ± 0.026**	0.81 ± 0.019	0.76 ± 0.022
RR	0.76 ± 0.016	0.74 ± 0.027	0.76 ± 0.046	0.75 ± 0.035	0.70 ± 0.046
SVM	0.77 ± 0.015	**0.93 ± 0.031**	0.72 ± 0.02	0.81 ± 0.015	0.74 ± 0.022
XGBoost	0.73 ± 0.035	0.84 ± 0.162	0.72 ± 0.037	0.77 ± 0.092	0.70 ± 0.083

Abbreviations: AUROC curve, area under the receiver operating characteristic curve; LR, logistic regression; ER, elastic-net regression; KNN, K-nearest neighbor; LASSO, least absolute shrinkage and selection operator; LDA, linear discriminant analysis; RF, random forest; RR, ridge regression; SVM, support vector machine.
The bold value indicates the highest performance in the column.

model simpler. The learning curve illustrated that, with 25 features being considered, the gap between the train and the validation error became steady as the sample size exceeded 200 (Figure 2B). The DF coefficient for each included feature is shown in Figure 2C. The absolute value of the coefficient reflected the contribution of the feature to the model. By summing up the product of the DF coefficient and the feature, a DF was calculated for each case. The distribution of the DF in each class was a bell-shaped curve, with a clearly distinct summit from each other (Figure 2D). As expected, the mean of the DF in each class was statistically different (Figure 2E). The plot of the DF (*x*-axis) over the probability of being predicted as positive by the final model (*y*-axis) demonstrated that, as the DF became greater, the probability of being predicted as positive declined (Figure 2F), which was consistent with the DF distribution across the classes (Figure 2E).

The SVM model with the tuned hyperparameters (Supplementary Table S2) performed best with respect to recall. Figure 3A shows that the sample size was sufficient to stabilize the validation error from the SVM model. Figures 3B,C

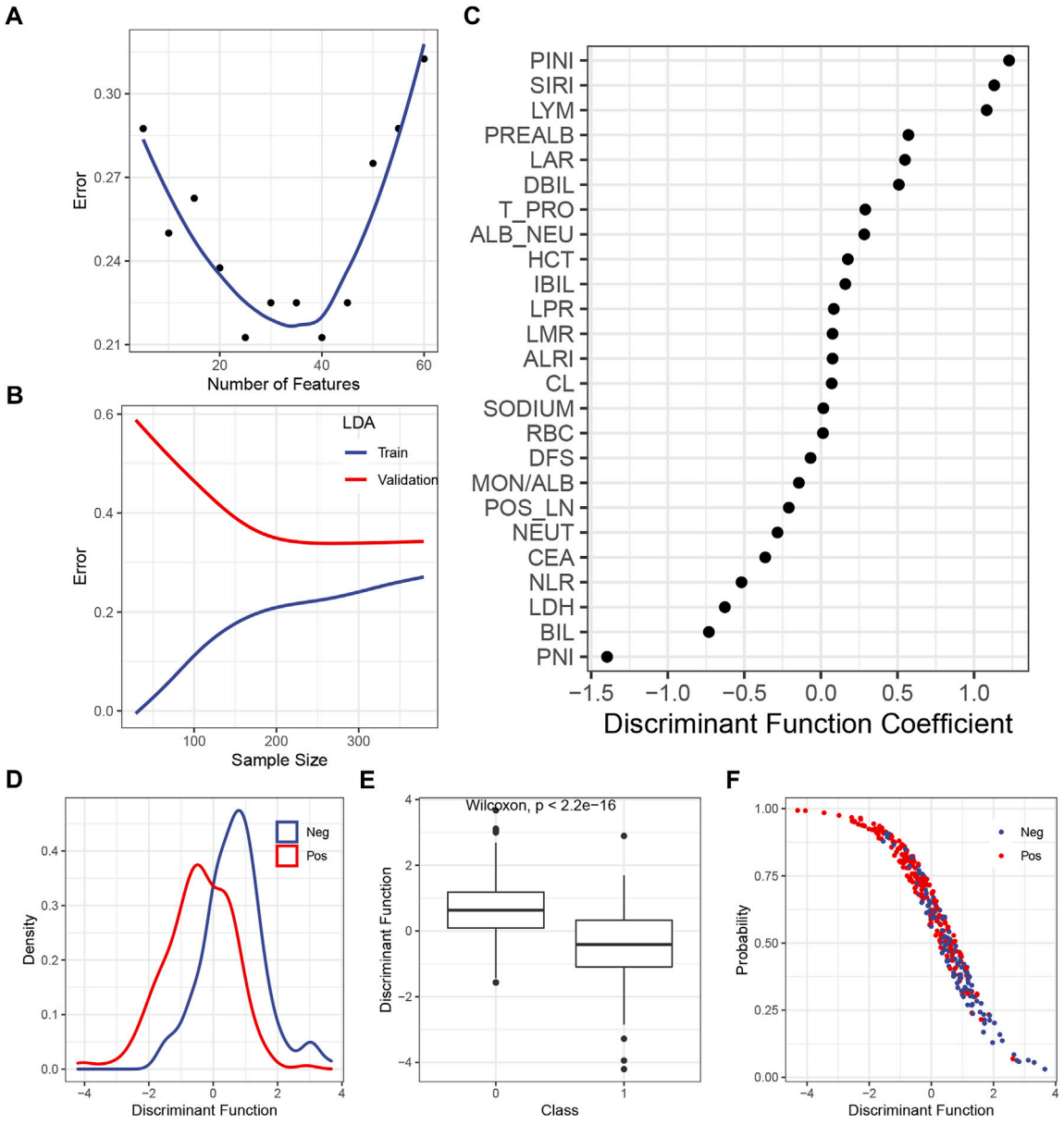


FIGURE 2
Construction of the LDA model. **(A)** Correlation between the predicted error in the validation set and the number of features. **(B)** Learning curve of the LDA model. **(C)** DF coefficient of the features in the model. All the features were scaled, so the absolute value of the coefficient reflected the contribution of a feature to the model. **(D)** Distribution of the DF in the positive or the negative subgroup. **(E)** Difference of the median DF in each subgroup was examined by the Wilcoxon test. **(F)** Plot of the DF over the predicted probability of being positive by the final LDA model. Abbreviations: LDA, linear discriminant analysis; DF, discriminant function; PINI, prognostic inflammatory nutrition index; SIRI, systemic inflammation response index; LYM, lymphocyte; PREALB, prealbumin; LAR, lymphocyte–albumin ratio; DIBL, direct bilirubin; T_PRO, total protein; NEU, neutrophil; HCT, hematocrit; IBIL, indirect bilirubin; LPR, lymphocyte platelet ratio; LMR, lymphocyte–monocyte ratio; ALRI, aspartate aminotransferase–lymphocyte ratio index; RBC, red blood cell; DFS, disease-free survival; MON, monocyte; ALB, albumin; POS_LN, number of positive lymph node; CEA, carcinoembryonic antigen; NLR, neutrophil–lymphocyte ratio; LDH, lactic dehydrogenase; BIL, bilirubin; PNI; prognostic nutrition index.

simulate hyperparameter tuning by grid search, showing that a combination of a radial *kernel* function, a natural log-transformed *cost* of about -3 and a natural log-transformed *gamma* of about -2, gave the highest AUROC in the validation set. It was consistent with the configuration of the actual model tuned by random search (Supplementary Table S2). Because this model only picked the first eight features, it is possible and more intuitive to interpret the SVM model by a

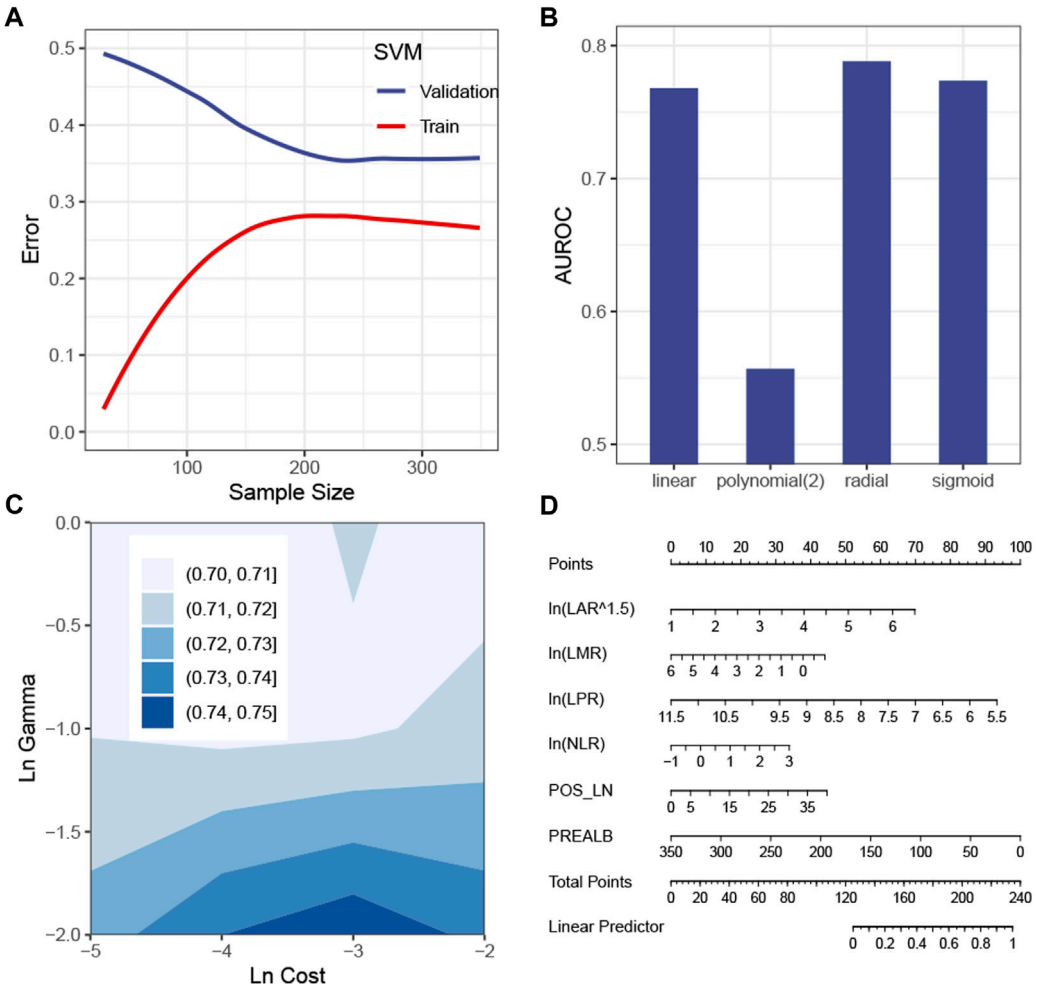


FIGURE 3
Construction of the SVM model. **(A)** Learning curve of the SVM model. **(B)** Simulation test of the tuning *kernel* function measured by the AUROC. **(C)** Simulation test of tuning *cost* and *gamma* measured by the AUROC. **(D)** Global surrogate model visualized by a nomogram to approximate the predicted probability of being positive by the SVM model. Abbreviations: SVM, support vector machine; AUROC curve, area under the receiver operating characteristic curve; LAR, lymphocyte–albumin ratio; LMR, lymphocyte–monocyte ratio; LPR, lymphocyte–platelet ratio; NLR, neutrophil–lymphocyte ratio; POS_LN, number of positive lymph node; PREALB, prealbumin.

linear regression model visualized by a nomogram (Figure 3D). Of the eight features, the linear model, namely, the surrogate model, automatically chose six features by a stepwise method (Figure 3D). The R-square was 0.68.

The RF model with the tuned hyperparameters (Supplementary Table S2) had the highest precision. Figure 4A shows that the model performance stabilized when the number of trees was greater than 400 and reached the best when approximately 700 trees were aggregated. The simulation test showed the optimal combination of the *mtry* and the *nodesize* by grid search (Figure 4B). The rank of the feature demonstrated that the baseline white cell count, platelet count, and prealbumin were the most crucial predictors in the RF model (Figure 4C). This model chose 29 features, so it was more appropriate to use a decision tree algorithm, with a *cp* of 0 and a *maxdepth* of 4, as a surrogate model to approximate the response predicted by the RF model (Figure 4D). The R-square was 0.57.

Discussion

This study demonstrates that the LDA, SVM, and RF models outperform other algorithms in predicting the survival of patients with metastatic GC. The model performance assessed by a single holdout CV is not reliable because of the effect of randomness in splitting data into training and validation sets. Here, this issue is addressed by repeating the holdout CV five

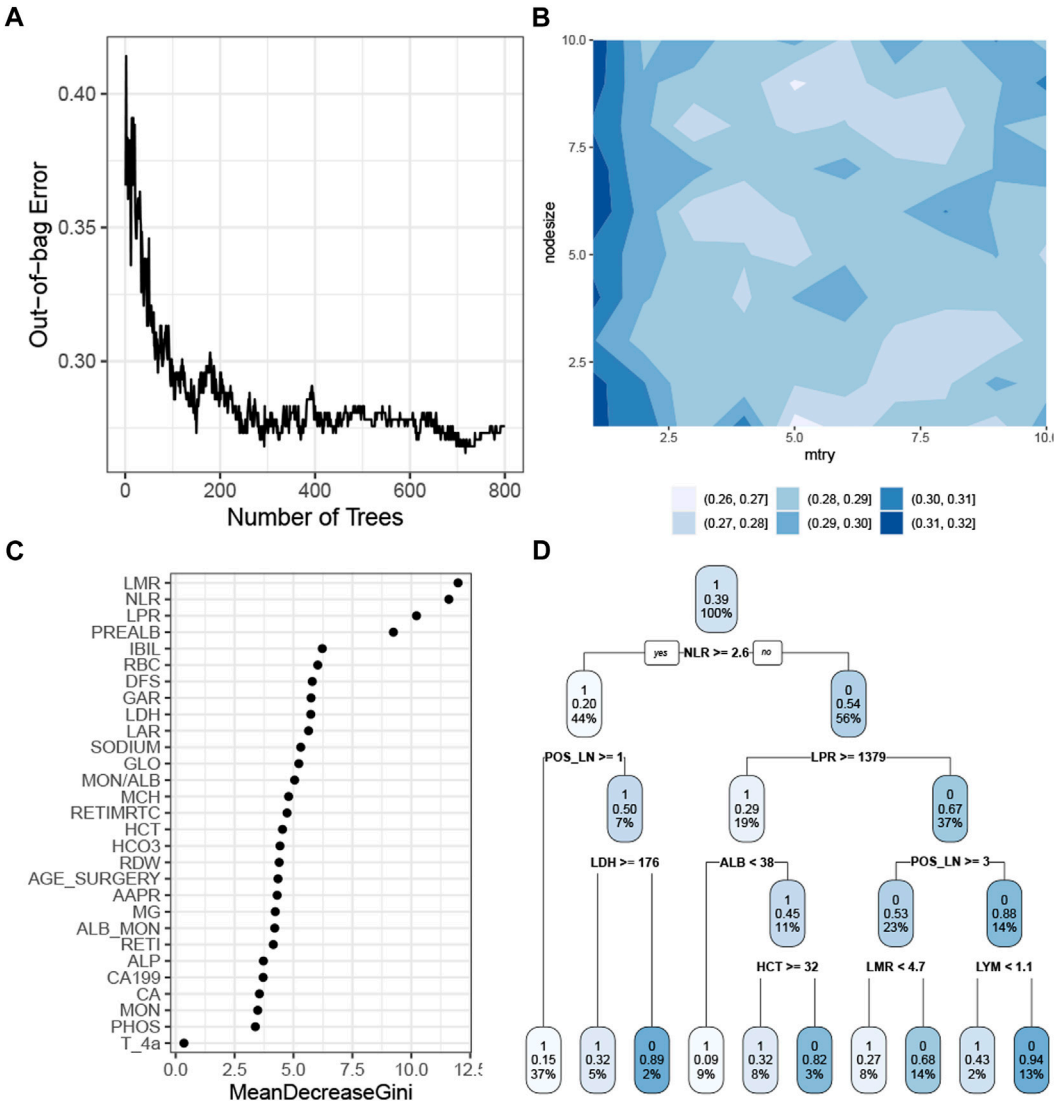


FIGURE 4
Construction of the RF model. **(A)** Correlation of the out-of-bag error with the number of trees. **(B)** Simulation test of tuning *mtry* and *nodesize* measured by the AUROC. **(C)** Feature importance as measured by the RF algorithm. **(D)** Global surrogate model visualized by a decision tree to approximate the predicted response by the RF model. Abbreviations: RF, random forest; AUROC curve, area under the receiver operating characteristic curve; LMR, lymphocyte–monocyte ratio; NLR, neutrophil–lymphocyte ratio; PREALB, prealbumin; IBIL, indirect bilirubin; RBC, red blood cell; DFS, disease-free survival; GAR, γ-glutamyl transpeptadase–albumin ratio; LDH, lactic dehydrogenase; LAR, lymphocyte–albumin ratio; GLO, globulin; MON, monocyte; ALB, albumin; MCH, mean corpuscular hemoglobin; RETIMRTC, moderate fluorescence–reticulocyte ratio; HCT, hematocrit; RDW, red cell distribution width; AAPR, albumin-to-alkaline phosphatase ratio; RETI, reticulocyte count; ALP, alkaline phosphatase; PHOS, phosphorus.

times, so the performance is unbiased in the local cohort, as measured by AUROC, recall, and precision. All the AUROCs of the three models exceed 0.75, which is higher than those in the previous reports (Koo et al., 2011; Custodio et al., 2017; Ma et al., 2021). In addition, our earlier work selected the patients with AGC receiving first-line chemotherapy from the same local cohort and compared the performances of seven survival prediction models on it, showing the best model having an AUROC of 0.60 (Xu et al., 2021). The present study does not impose restrictions on treatment or other factors, so the population is of wider heterogeneity, which may help the model generalize well due to closely reflecting the characteristics of the patients in a real-world setting.

The LDA model performs the best as measured by the AUROC. This model is relatively the most interpretable model among the algorithms tested because it constructs a single and intuitive DF from the original features for classification, which is similar to the well-known logistic

regression. In contrast, the SVM and RF models are less explainable; however, they perform the best as measured by recall and precision, respectively. In clinical practice, if we want to identify as many patients as possible with less than 1-year survival by the model, i.e., avoid a false-negative event, we should consider the SVM model because it has the highest recall. If we want to be confident with the ML's prediction, i.e., avoid a false-positive event, we should consider the RF model due to its highest precision. Therefore, model selection is closely related to the specific practical scenario.

Almost all the features included for modeling are the baseline index of systemic inflammation and malnutrition, albeit with a few exceptions, which is in line with the previous studies. For patients receiving first-line treatment, Kim et al. (2020) identified the clinically relevant features as NLR, neutrophil count, alkaline phosphatase, albumin, lymphocyte count, and white blood cell count. The AGAMENON nomogram weights the neutrophil-to-lymphocyte ratio (NLR) greater than tumor differentiation, metastasis site, or HER2+-treated (Custodio et al., 2017). Hsieh et al. (2016) selected NLR, modified Glasgow prognostic score (mGPS), and Patient-Generated Subjective Global Assessment (PG-SGA) as the most relevant predictors (all of them are inflammation- or nutrition-based scores), compared with age, physical status, differentiation, and metastasis site. In fact, the laboratory index is frequently adopted in models for predicting the survival of metastatic GC patients (Chau et al., 2004; Lee et al., 2007; Kim et al., 2008; Takahari et al., 2014; Wang et al., 2016; Kim et al., 2020), as extra information added on the well-recognized predictors. The present study further emphasizes the prediction value of the index in a setting of a more heterogeneous population with GC. In other words, the ML model "considers" that the post-metastasis survival is mainly attributed to the characteristics of the patients at the time of metastasis, other than the history of staging, surgery, adjuvant chemotherapy, and so on. This may explain the prognoses of the patients with metachronous or synchronous GC being the same as long as the cancer is at an advanced stage (Patel et al., 2007).

The current study has several limitations. First, the models are confident to generalize well in the local population because they are unbiasedly evaluated by repeated CV; however, the case is uncertain in other situations due to lack of external validation. Second, the ensemble technique, a common method to enhance model performance (Rhys, 2020), is not utilized for modeling. We consider there is a trade-off between model complexity and model performance: using ensemble is very likely to improve the model at the cost of long running time and poor interpretability and vice versa. So, here we prefer an easier model, at the cost of performance, in order to facilitate real clinical practice. Third, there are still over 20 features in some models, albeit feature filtering has been conducted. This would impede the models from being used as a quick screening tool for practitioners.

In conclusion, on the basis of the readily available information from the EMR and LIS, the mainstream ML method can produce satisfactory models for predicting survival in patients with metastatic GC who experienced prior radical gastrectomy. The algorithm should be selected according to the measurement and its meaning in a practical scenario.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of The First Affiliated Hospital of Anhui Medical University. The Ethics Committee waived the requirement of written informed consent for participation.

Author contributions

Conceptualization: CF and TM; data curation: YZ, M-MX, X-PZ, CF, Z-JW, YF, and Y-HY; formal analysis: CZ, YZ, and Y-HY; funding acquisition: TM; investigation: M-MX, YZ, X-PZ, CF, Z-JW, YF, and Y-HY; methodology: CF and TM; project administration: CF and TM; resources: M-MX, YZ, X-PZ, CF, Z-JW, YF, and Y-HY; supervision: CF and TM; validation: M-MX, YZ, X-PZ, CF, Z-JW, YF, and Y-HY; writing—original draft: CZ, YZ, and Y-HY; writing—review and editing: CZ, HX, TM, M-MX, YZ, X-PZ, CF, Z-JW, YF, and Y-HY.

Funding

This work was supported by the Anhui Provincial Key Research and Development Program [grant number 1804b06020351], The First Affiliated Hospital of Anhui Medical University Clinical Research Project [grant number LCYJ2021YB015], and the Research Fund from Anhui Medical University [grant number 2022xkj160].

Acknowledgments

The authors are grateful to Mr. Zhen-Hui He for his help in data electronization.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.937242/full#supplementary-material

References

Boehmke, B., and Greenwell, B. (2019). *Hands-on machine learning with R*. Boca Raton, Florida, United States: CRC Press.

Chau, I., Norman, A. R., Cunningham, D., Waters, J. S., Oates, J., and Ross, P. J. (2004). Multivariate prognostic factor analysis in locally advanced and metastatic esophago-gastric cancer-pooled analysis from three multicenter, randomized, controlled trials using individual patient data. *J. Clin. Oncol.* 22 (12), 2395–2403. doi:10.1200/JCO.2004.08.154

Chen, M., He, F. Q., Liao, M. S., Yang, C., and Chen, X. D. (2021). Gastrectomy with omentum preservation versus gastrectomy with omentectomy for locally advanced gastric cancer: A systematic review and meta-analysis. *Int. J. Surg.* 96, 106176. doi:10.1016/j.ijsu.2021.106176

Custodio, A., Carmona-Bayonas, A., Jimenez-Fonseca, P., Sanchez, M. L., Viudez, A., Hernandez, R., et al. (2017). Nomogram-based prediction of survival in patients with advanced oesophagogastric adenocarcinoma receiving first-line chemotherapy: A multicenter prospective study in the era of trastuzumab. *Br. J. Cancer* 116 (12), 1526–1535. doi:10.1038/bjc.2017.122

Elshawi, R., Al-Mallah, M. H., and Sakr, S. (2019). On the interpretability of machine learning-based model for predicting hypertension. *BMC Med. Inf. Decis. Mak.* 19 (1), 146. doi:10.1186/s12911-019-0874-0

He, H., and Ma, Y. (2013). *Imbalanced learning. Foundations, algorithms, and applications*. Hoboken, New Jersey, United States: Wiley-IEEE Press.

Hisamori, S., Okabe, H., Tsunoda, S., Nishigori, T., Ganeko, R., Fukui, Y., et al. (2021). Long-term outcomes of laparoscopic radical gastrectomy for highly advanced gastric cancer: Final report of a prospective phase II trial (KUGC04). *Ann. Surg. Oncol.* 28 (13), 8962–8972. doi:10.1245/s10434-021-10373-8

Hsieh, M. C., Wang, S. H., Chuah, S. K., Lin, Y. H., Lan, J., and Rau, K. M. (2016). A prognostic model using inflammation- and nutrition-based scores in patients with metastatic gastric adenocarcinoma treated with chemotherapy. *Med. Baltim.* 95 (17), e3504. doi:10.1097/MD.0000000000003504

Kim, J., Hong, J. Y., Kim, S. T., Park, S. H., Jekal, S. Y., Choi, J. S., et al. (2020). Clinical scoring system for the prediction of survival of patients with advanced gastric cancer. *ESMO Open* 5 (2), e000670. doi:10.1136/esmoopen-2020-000670

Kim, J. G., Ryoo, B. Y., Park, Y. H., Kim, B. S., Kim, T. Y., Im, Y. H., et al. (2008). Prognostic factors for survival of patients with advanced gastric cancer treated with cisplatin-based chemotherapy. *Cancer Chemother. Pharmacol.* 61 (2), 301–307. doi:10.1007/s00280-007-0476-x

Koo, D. H., Ryoo, B. Y., Kim, H. J., Ryu, M. H., Lee, S. S., Moon, J. H., et al. (2011). A prognostic model in patients who receive chemotherapy for metastatic or recurrent gastric cancer: Validation and comparison with previous models. *Cancer Chemother. Pharmacol.* 68 (4), 913–921. doi:10.1007/s00280-011-1561-8

Lee, J., Lim, T., Uhm, J. E., Park, K. W., Park, S. H., Lee, S. C., et al. (2007). Prognostic model to predict survival following first-line chemotherapy in patients with metastatic gastric adenocarcinoma. *Ann. Oncol.* 18 (5), 886–891. doi:10.1093/annonc/mdl501

Ma, T., Wu, Z., Zhang, X., Xu, H., Feng, Y., Zhang, C., et al. (2021). Development and validation of a prognostic scoring model for mortality risk stratification in patients with recurrent or metastatic gastric carcinoma. *Bmc Cancer* 21 (1), 1326. doi:10.1186/s12885-021-09079-7

Patel, P. R., Yao, J. C., Hess, K., Schnirer, I., Rashid, A., and Ajani, J. A. (2007). Effect of timing of metastasis/disease recurrence and histologic differentiation on survival of patients with advanced gastric cancer. *Cancer* 110 (10), 2186–2190. doi:10.1002/cncr.23046

Rhys, H. I. (2020). *Machine Learning with R, the tidyverse, and mlr*. Shelter Island, New York, United States: Manning.

Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C., and Lordick, F. (2020). Gastric cancer. *Lancet* 396 (10251), 635–648. doi:10.1016/S0140-6736(20)31288-5

Takahari, D., Boku, N., Mizusawa, J., Takashima, A., Yamada, Y., Yoshino, T., et al. (2014). Determination of prognostic factors in Japanese patients with advanced gastric cancer using the data from a randomized controlled trial, Japan clinical oncology group 9912. *Oncologist* 19 (4), 358–366. doi:10.1634/theoncologist.2013-0306

Wang, J., Qu, J., Li, Z., Che, X., Zhang, J., Liu, J., et al. (2016). A prognostic model in metastatic or recurrent gastric cancer patients with good performance status who received first-line chemotherapy. *Transl. Oncol.* 9 (3), 256–261. doi:10.1016/j.tranon.2016.04.004

Xu, H., Zhang, X., Wu, Z., Feng, Y., Zhang, C., Xie, M., et al. (2021). Performances of prognostic models in stratifying patients with advanced gastric cancer receiving first-line chemotherapy: A validation study in a Chinese cohort. *J. Gastric Cancer* 21 (3), 268–278. doi:10.5230/jgc.2021.21.e26

Zhu, X. D., Huang, M. Z., Wang, Y. S., Feng, W. J., Chen, Z. Y., He, Y. F., et al. (2022). XELOX doublet regimen versus EOX triplet regimen as first-line treatment for advanced gastric cancer: An open-labeled, multicenter, randomized, prospective phase III trial (EXELOX). *Cancer Commun.* 42 (4), 314–326. doi:10.1002/cac2.12278

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 20 December 2022
DOI 10.3389/fmolb.2022.991425

Check for updates

OPEN ACCESS

EDITED BY
Wei Wang,
First Affiliated Hospital of Anhui Medical University, China

REVIEWED BY
Robert H. Newman,
North Carolina Agricultural and Technical State University, United States
Qian Xiaoqin,
Jiangsu University Affiliated People's Hospital, China

*CORRESPONDENCE
Ling Weng,
lingweng@csu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 11 July 2022
ACCEPTED 05 December 2022
PUBLISHED 20 December 2022

CITATION
Li S, Han Y, Zhang Q, Tang D, Li J and Weng L (2022), Comprehensive molecular analyses of an autoimmune-related gene predictive model and immune infiltrations using machine learning methods in moyamoya disease. *Front. Mol. Biosci.* 9:991425. doi: 10.3389/fmolb.2022.991425

COPYRIGHT
© 2022 Li, Han, Zhang, Tang, Li and Weng. 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.

Comprehensive molecular analyses of an autoimmune-related gene predictive model and immune infiltrations using machine learning methods in moyamoya disease

Shifu Li[1,2†], Ying Han[3,4†], Qian Zhang[1,2], Dong Tang[1,2], Jian Li[1,2,5] and Ling Weng[2,6,7,8,9*]

[1]Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]National Clinical Research Center for Geriatric Disorders, Central South University, Changsha, Hunan, China, [3]Department of Oral and Maxillofacial Surgery, Center of Stomatology, Xiangya Hospital of Central South University, Changsha, Hunan, China, [4]Center for Medical Genetics and Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China, [5]Hydrocephalus Center, Xiangya Hospital, Central South University, Changsha, Hunan, China, [6]Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan, China, [7]Engineering Research Center of Hunan Province in Cognitive Impairment Disorders, Central South University, Changsha, China, [8]Hunan International Scientific and Technological Cooperation Base of Neurodegenerative and Neurogenetic Diseases, Changsha, China, [9]Key Laboratory of Hunan Province in Neurodegenerative Disorders, Central South University, Changsha, China

Background: Growing evidence suggests the links between moyamoya disease (MMD) and autoimmune diseases. However, the molecular mechanism from genetic perspective remains unclear. This study aims to clarify the potential roles of autoimmune-related genes (ARGs) in the pathogenesis of MMD.

Methods: Two transcription profiles (GSE157628 and GSE141025) of MMD were downloaded from GEO databases. ARGs were obtained from the Gene and Autoimmune Disease Association Database (GAAD) and DisGeNET databases. Differentially expressed ARGs (DEARGs) were identified using "limma" R packages. GO, KEGG, GSVA, and GSEA analyses were conducted to elucidate the underlying molecular function. There machine learning methods (LASSO logistic regression, random forest (RF), support vector machine-recursive feature elimination (SVM-RFE)) were used to screen out important genes. An artificial neural network was applied to construct an autoimmune-related signature predictive model of MMD. The immune characteristics, including immune cell infiltration, immune responses, and HLA gene expression in MMD, were explored using ssGSEA. The miRNA-gene regulatory network and the potential therapeutic drugs for hub genes were predicted.

Results: A total of 260 DEARGs were identified in GSE157628 dataset. These genes were involved in immune-related pathways, infectious diseases, and autoimmune diseases. We identified six diagnostic genes by overlapping the three machine learning algorithms: CD38, PTPN11, NOTCH1, TLR7, KAT2B, and

ISG15. A predictive neural network model was constructed based on the six genes and presented with great diagnostic ability with area under the curve (AUC) = 1 in the GSE157628 dataset and further validated by GSE141025 dataset. Immune infiltration analysis showed that the abundance of eosinophils, natural killer T (NKT) cells, Th2 cells were significant different between MMD and controls. The expression levels of HLA-A, HLA-B, HLA-C, HLA-DMA, HLA-DRB6, HLA-F, and HLA-G were significantly upregulated in MMD. Four miRNAs (mir-26a-5p, mir-1343-3p, mir-129-2-3p, and mir-124-3p) were identified because of their interaction at least with four hub DEARGs.

Conclusion: Machine learning was used to develop a reliable predictive model for the diagnosis of MMD based on ARGs. The uncovered immune infiltration and gene-miRNA and gene-drugs regulatory network may provide new insight into the pathogenesis and treatment of MMD.

KEYWORDS

moyamoya disease, machine learning, bioinformatics, immune infiltration, autoimmune-related genes

Introduction

Moyamoya disease (MMD) is an uncommon, chronic cerebrovascular disorder characterized by progressive occlusion of the supraclinoid internal carotid artery (ICA) and its main branches within the circle of Willis. MMD, also known as an abnormal netlike vascular disease at the base of the brain, is a term coined by Suzuki and Takaku in 1969 to describe the classic angiographic appearance—a puff of cigarette smoke drifting in the air (Suzuki and Takaku, 1969). Important clinical features include ischemic stroke, often presented in childhood, and hemorrhagic stroke, generally observed in adults. The incidence rate is twice higher in females than males, and children around 5 years and adults in their mid-40s are particularly affected (Kuroda and Houkin, 2008). MMD is the most common pediatric cerebrovascular disease in Japan, affecting approximately three out of every 100,000 children (Scott and Smith, 2009). MMD conventionally refers to patients with above idiopathic pathology without a previously diagnosed condition. Distinct from the definitive MMD, Moyamoya (MM) syndrome (also named Quasi-Moyamoya disease, secondary Moyamoya disease, and akin-Moyamoya disease) is the occurrence of angiographic MM in association with acquired (i.e., autoimmune diseases) or inherited disorders [i.e., neurofibromatosis type 1, sickle cell anemia, Down syndrome (DS)] (Scott and Smith, 2009).

The exact etiology of MMD remains unknown; however, hereditary, immunogenic inflammatory, and hemodynamic factors are known to be responsible. The close relationship between patients with MM syndrome and autoimmune diseases, such as type 1 diabetes mellitus, thyroid disease, systemic lupus erythematosus (SLE), and DS has been reported (Huang et al., 2017). A study in a primarily white, midwestern United States population showed that the prevalence of autoimmune diseases was significantly higher in patients with MMD, particularly type 1 diabetes mellitus (8.5% *versus* 0.4% in the general population), thyroid disease (17.0% *versus* 8.0% in the institutional general patient population), and hyperlipidemia (27.7% *versus* 16.3% in the general population). A meta-analysis conducted in 2014 revealed that elevated thyroid autoantibodies and elevated thyroid function are independently associated risk factors for MMD (Lei et al., 2014). Autoimmunity is the main link between SLE and MM syndrome since immune complexes lead to vasculitis and narrow or occluded vessels (El Ramahi and Al Rayes, 2000; Jeong et al., 2008). Chen et al. (2016) found that the overall prevalence of autoimmune diseases in patients with unilateral MMD was significantly higher than that in patients with bilateral MMD. Although the close relationships between autoimmune diseases and MDD diseases have been recognized, the underlying mechanisms remain to be clarified.

Advances in molecular biology and next-generation sequencing technologies have made it possible to study disease mechanisms at the genetic and mRNA levels. Gene expression profiling through methods such as microarray and RNA sequencing based on the Gene Expression Omnibus (GEO) database is widely used to explore differentially expressed genes (DEGs), analyze potential function pathways, and determine molecular mechanisms involved in various cerebrovascular diseases (Chen et al., 2022). A recent bioinformatics study identified the potential neutrophil-associated genes in MMD (Jin and Duan, 2022). However, the role of autoimmune-related genes (ARGs) in the pathophysiology of MMD is still unclear. In recent years, the development of machine learning algorithms has provided more choices for diagnostic models as precision medical predictive tools. Our study integrated least absolute shrinkage and selection operator (LASSO) logistic regression, random forest (RF), support vector machine-recursive feature elimination (SVM-RFE), and artificial neural network to screen and identify

diagnostic markers and construct an autoimmune-related signature predictive model of MMD. The immune characteristics, including immune cell infiltration, immune responses, and HLA gene expression in MMD, were explored. The miRNA-gene regulatory network and the potential therapeutic drugs for hub genes were predicted.

Materials and methods

Downloading and processing of data

Microarray data containing two transcription profiles (GSE157628 and GSE141025) were downloaded from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/). The dataset of GSE157628 was utilized as the exploratory dataset, and the GSE141025 profile acted as the validation dataset. The GSE157628 profile included micro-samples of the middle cerebral artery (MCA) collected from 11 patients with MMD and nine age- and gender-matched control samples (three from patients with epilepsy and six from patients with ICA aneurysms) at the platform of GPL16699. The expression profiles of MCA samples from four patients with MMD and four matched superficial temporal artery controls in GSE141025 were extracted for dataset validation. If multiple probes matched one gene, the probe with the maximal median expression values was annotated into the homologous gene symbol through the platform's annotation information.

Collection of autoimmune-related genes

ARGs were obtained from the Gene and Autoimmune Disease Association Database (GAAD) (Lu et al., 2018) and DisGeNET databases (Piñero et al., 2020) after deleting duplicate genes. GAAD contained 44,762 associations between 49 autoimmune-related diseases and 4,249 genes through text mining and manual curation. DisGeNET (v7.0), one of the largest publicly available collections of genes and variants associated with human diseases, contained 1,134,942 gene-disease associations (GDAs), between 21,671 genes and 30,170 diseases, disorders, traits, and clinical or abnormal human phenotypes.

Identification of differentially expressed autoimmune-related genes

The principal component analysis was conducted by using the factoextra R package. To identify the DEARGs, we performed differential expression analysis using the "limma" package in R software to detect DEGs between the MMD and control groups in the training dataset (GSE157628). The DEGs were screened with the criteria of $|\log2FoldChange| > 1$ and $p < 0.05$. Volcano maps and clustering heatmaps were prepared to visualize the differences using the "ggplot2" and "ComplexHeatmap" packages in the R software. We intersected the DEGs with ARGs to identify DEARGs and visualized them with the "VennDiagram" package.

Functional enrichments between moyamoya disease tissues and controls

To uncover the biological function in MMD, gene set enrichment analysis (GSEA) was used to enrich the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the "clusterProfiler" package. Gene set variation analysis (GSVA) algorithm was used to calculate the Reactome ("c2.cp.reactome.v7.5.1.symbols" gene set from the Molecular Signatures Database) (Liberzon et al., 2015) processes score using the "GSVA" R package.

Authentication of the organ/tissue-specific expressed DEARGs

To understand the tissue/organ-specific expression of these DEARGs, we analyzed the gene distribution in tissues using the online tool BioGPS (http://biogps.org/) (Wu et al., 2009). The following criteria had to be met (Wang et al., 2020): 1) the expression level of transcripts mapped to a single organ system was >10 times the median, and 2) the second-highest level was not more than one-third of the highest expression level.

Protein-protein interaction network and functional annotation of DEARGs

The PPI network of DEARGs was prepared using the online tool STRING (https://string-db.org/) with a minimum required interaction score of 0.4. We downloaded the interaction information and visualized the PPI network using Cytoscape software (v3.8.2). ClueGO, a plugin app of Cytoscape for the function enrichment, was used to annotate the biological processes (BP) of Gene Ontology (GO) and KEGG pathways of this network and the genes participating in these terms. The R package "DOSE" was used to perform Disease Ontology (DO) analysis (Yu et al., 2015).

Screening for crucial DEARGs and candidate signatures

First, we applied five methods (Closeness, Degree, MCC(Maximal Clique Centrality), MNC (Maximum

neighborhood component), and Radiality) in cytoHubba to select the top 30 genes and intersected them through the Venn plot to find the common genes (Chin et al., 2014). Based on common genes, we constructed a co-expression network *via* GeneMANIA (http://www.genemania.org/) to identify internal associations (Warde-Farley et al., 2010).

Second, to further identify the crucial DEARGs and candidate signatures, three machine learning [LASSO logistic regression, RF, and support vector machine (SVM)] algorithms were adopted. The LASSO logistic regression model was used to select optimal variables using the penalty coefficient. RF is a machine learning algorithm with an ensemble of multiple decision trees that combines the knowledge generated by a collection of individual trees using randomness. The top 10 variables were selected as the most important features in the methods. SVM is a supervised machine learning technique widely utilized for classification and regression. To avoid over-fitting, the SVM-RFE requires training multiple classifiers on subsets of features of decreasing size to search for the best features. In the present study, we overlapped genes identified by the three methods and designated the communal gens as our candidate signatures for constructing the next diagnostic model.

Construction of the artificial neural network diagnostic model

We constructed a back propagation artificial neural network model using the "neuralnet" package. The expression profiles of the above screened signatures in the GSE157628 dataset were extracted and normalized. The min-max method was selected, and the data were mapped in the range of zero to one before training the neural network. The number of neurons was between the input and output layer sizes, usually two-thirds of the input size. A single hidden layer with four nodes was used. We calculated the classification score of the obtained disease neural network model as follows:

$$\text{Predicting score} = \sum gene\ expression^{*}neural\ network\ weight$$

The diagnostic ability was evaluated through the receiver operating characteristic (ROC) curve and confusion matrix. Another external dataset, GSE141025, was used to validate this model.

Immune characteristics of the moyamoya disease microenvironment

xCell Aran et al. (2017), a novel gene signature-based method to identify 64 immune and stromal cell types, was used to score the abundance of immune cells in MMD. Single sample gene set enrichment analysis (ssGSEA) was used to analyze the immune cells and activities between MMD and controls. We also compared the expression levels of HLA molecules between the two groups. The significantly different immune characteristics were depicted with boxplots and heatmaps.

The association of the crucial gene signatures with the scores of infiltrating immune cells, immune activities, and expression of HLA molecules was explored using Pearson's correlation analysis in R software. The resulting associations were visualized as a heatmap prepared with the "ggplot2" package.

Prediction of a miRNA-genes regulatory network and potential drugs

NetworkAnalyst is a user-friendly online tool to create PPI networks, cell-type or tissue-specific PPI networks, gene regulatory networks, gene co-expression networks, and networks for toxicogenomics and pharmacogenomics studies (Zhou et al., 2019). We used the NetworkAnalyst to predict the miRNA-genes regulatory network through the Tarbase database (Karagkouni et al., 2018).

The Drug-Gene Interaction Database (DGIdb) (http://www.dgidb.org/) is an online database to predict drug-gene interaction based on the data mined from DrugBank, PharmGKB, Chembl, Drug Target Commons, and TTD. The DGIdb was searched to make predictions on potential molecule-related drugs that interact with crucial DEARGs. Only the drugs with identified interaction types persisted, and the Drugs-Genes interactions were visualized through a Sankey diagram.

Sample collection and real-time quantitative polymerase chain reaction

We recruited ten patients who were diagnosed with MMD in Xiangya Hospital of Central South University for this research between June 2022 and October 2022. A total of 10 healthy controls with gender and age matched were also selected.

The detailed procedures of RT-qPCR were described in our previous studies (Li et al., 2022a; Li et al., 2022b). Birefly, peripheral blood monocytes (PBMCs) were isolated from the blood samples of patients and health persons. We extracted the total RNA from the PBMCs, performed reverse transcription reactions, and then amplified the cDNA. The results were analyzed using the $2^{-\Delta\Delta CT}$ method and expressed as ratio of the internal control, GAPDH. The primer sequences used for RT-qPCR are listed in Supplementary Table S1.

Results

Differential expression analysis

The study flowchart is depicted in Figure 1. Principal component analysis showed that the MMD tissues and

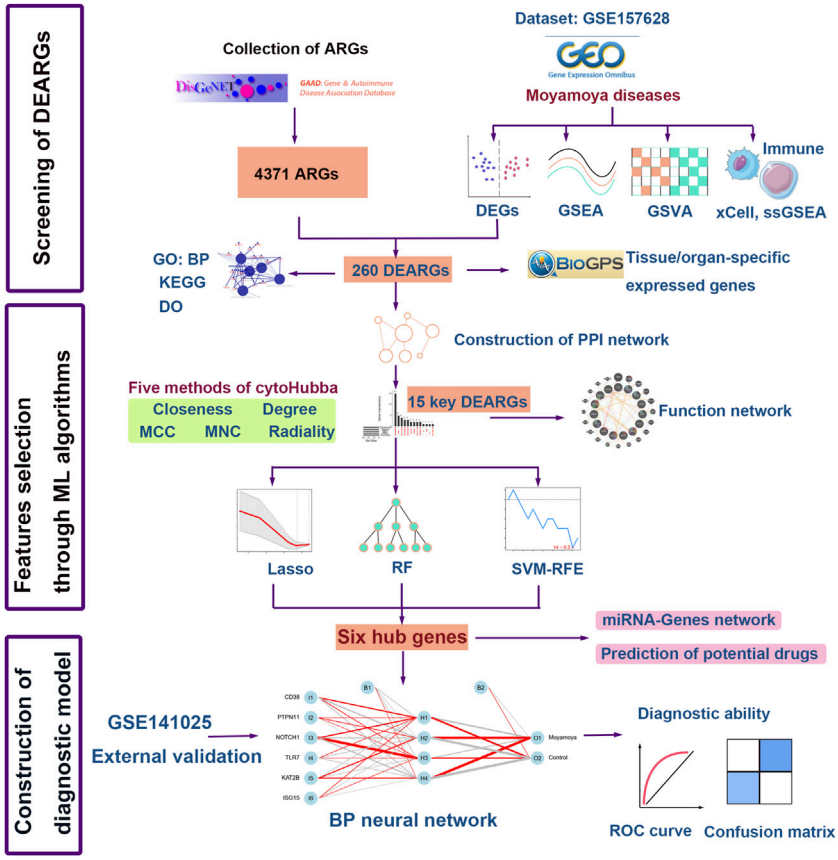


FIGURE 1
Flow chart of the study. DEGs, differentially expressed genes; GSEA, gene set enrichment analysis; ssGSEA, Single sample gene set enrichment; GSVA, gene set variation analysis; GO: BP, Gene Ontology: biological processes; KEGG, Kyoto Encyclopedia of Genes and Genomes; DO, Disease Ontology; ARGs, autoimmune-related genes; DEARGs, differentially expressed autoimmune-related genes; LASSO, the least absolute shrinkage and selection operation; RF, random forest; SVM-RFE, Support vector machine-recursive feature elimination; ML, machine learning; BP neural network, back propagation neural network; ROC, Receiver operating characteristic curve.

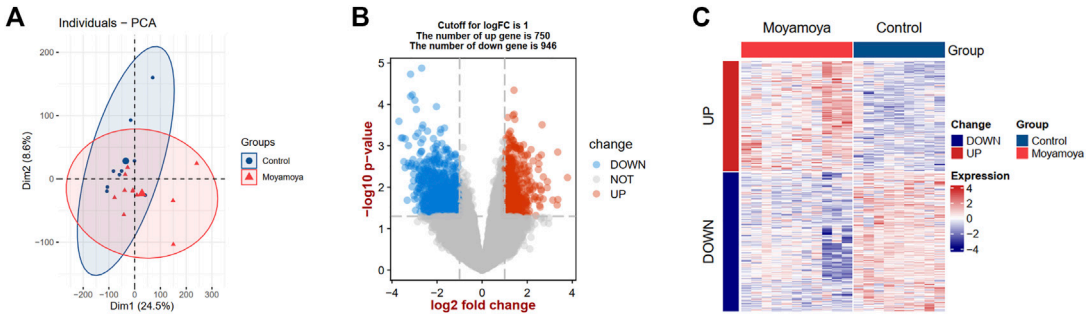


FIGURE 2
PCA and DEG analysis between MMD tissues and controls. **(A)** Principal component analysis between MMD tissues and controls. **(B)** A volcano plot shows the DEGs. Blue dots show the down-regulated genes and red dots represent the up-regulated genes. **(C)** A heat map shows the expression patterns of DEGs. PCA, principal component analysis; DEGs, differentially expressed genes; MMD, moyamoya disease.

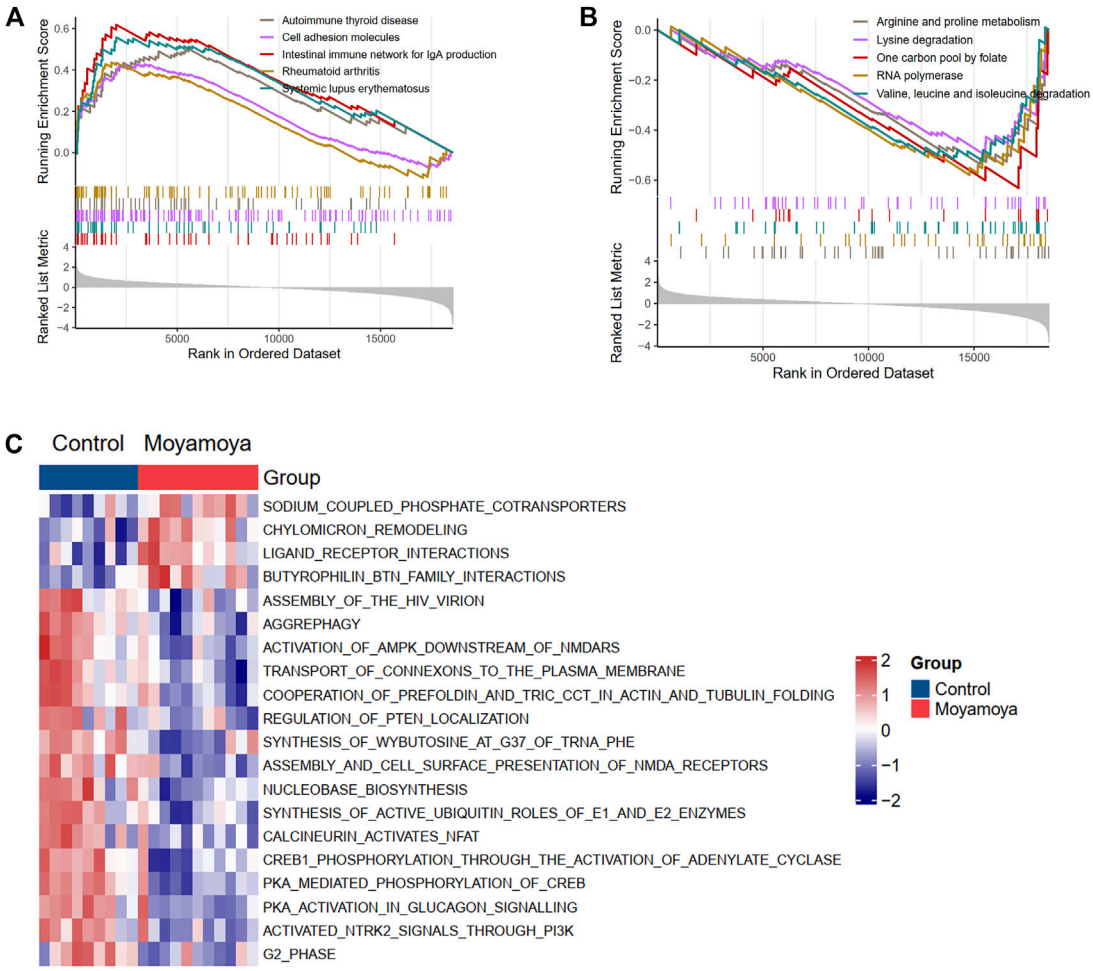


FIGURE 3
Biological KEGG and Reactome pathways involved in MMD based on GSEA and GSVA. **(A)** and **(B)** Up-regulated and down-regulated KEGG pathways from GSEA results, respectively. **(C)** The top differentially regulated reactome pathways from GSVA results.

controls could be clearly distinguished in the GSE157628 dataset (Figure 2A). Differential expression analysis was further performed to screen for DEGs. Based on the selection criteria, 1696 DEGs (750 upregulated and 946 downregulated) were identified (Figure 2B). The expression patterns of these DEGs were visualized through a hierarchical clustering heatmap (Figure 2C).

Functional enrichment analyses in moyamoya disease

GSEA and GSVA were performed to reveal the underlying biological pathways in MMD. The KEGG and Reactome analysis focus on the biological pathways. Compared with the comprehensive KEGG pathway, the Reactome pathway has more specific functions and focuses more on biochemical reactions. In our analysis, both methods were adopted. The KEGG analysis based on GSEA showed that the autoimmune thyroid disease, cell adhesion molecules, and rheumatoid arthritis pathways were upregulated (Figure 3A). In contrast, metabolism-related pathways (arginine and proline metabolism, lysine degradation, and one carbon pool by folate) were downregulated in MMD tissues (Figure 3B) compared with controls. The top 20 significantly differential Reactome pathways between MMD and controls are presented in Figure 3C, showing upregulated sodium-coupled phosphate cotransporters, chylomicron remodeling, and ligand-receptor interactions pathways. At the same time, aggrephagy and regulation of PTEN localization were downregulated in MMD tissues compared with controls (Figure 3C).

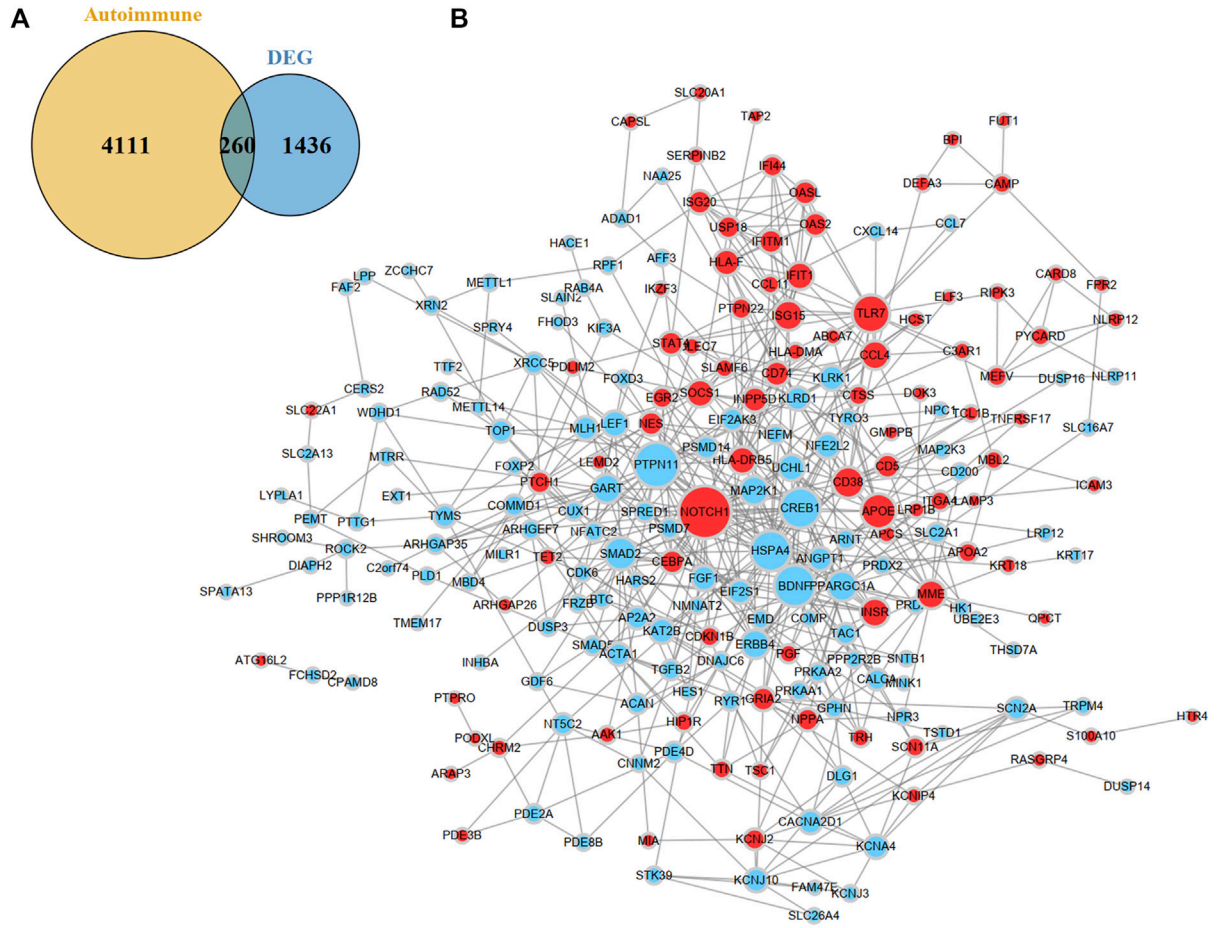


FIGURE 4
Identification and PPI network construction of DEARGs. **(A)** A venn plot show 260 DEARGs in MMD. **(B)** PPI network of DEARGs. The blue nodes represent the down-regulated genes and the red nodes represent the up-regulated genes. The dot size indicates the degree of the nodes. DEARGs, differentially expressed autoimmune-related genes.

Construction of protein-protein interaction network of DEARGs

After combining the GAAD and DisGeNET databases, 4371 ARGs were obtained. We overlapped the DEGs and ARGs, resulting in 260 DEARGs in MMD (Figure 4A). To elucidate the molecules' functional associations, we imported these genes into the STRING database to construct a PPI network and further visualized it in Cytoscape. After removing nodes without interaction with other genes, a PPI network with 225 nodes and 602 edges was constructed (Figure 4B).

Functional enrichment analyses of DEARGs

For exploring the function and pathway of the DEARGs, we used ClueGO, a plugin of Cytoscape, and the "DOSE" package. The biological processes-genes network showed that mononuclear cell migration and regulation of viral life cycle were most enriched for DEARGs (Figure 5A and Supplementary Figure S1A). The cAMP signaling pathway, PI3K-Akt signaling pathway, antigen processing and presentation, and microbial infection-related pathways were enriched in the KEGG analysis (Figure 5B and Supplementary Figure S1B). To uncover the diseases that these genes may be involved in, a DO analysis was conducted, which showed that these genes participated in the development of different cancers, infectious diseases, and autoimmune diseases.

Identification of tissue/organ-specific expressed genes

A total of 70 tissue/organ-specific expressed genes were identified for 260 DEARGs by BioGPS (Table 1). We observed

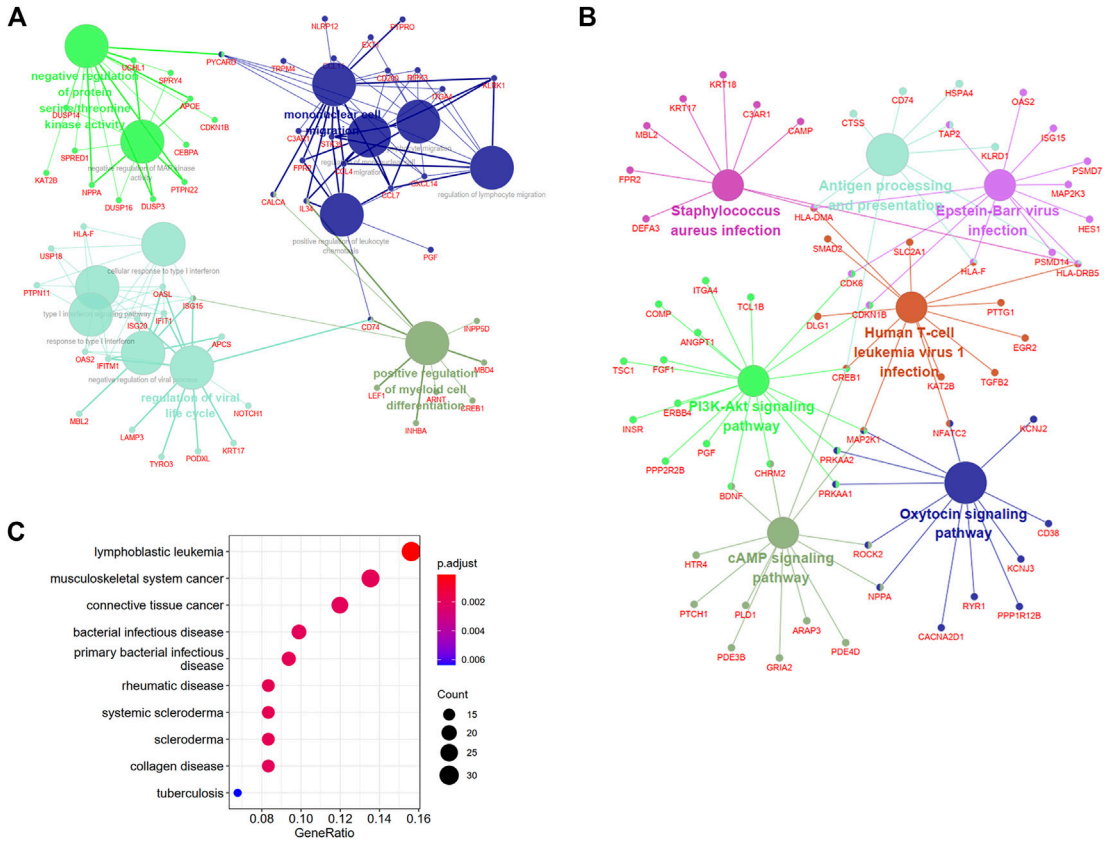


FIGURE 5
Biological annotations and DO analysis of DEARGs. **(A)** and **(B)** The enrichment network of biological processes and KEGG pathways with the participated genes in Cluego software. **(C)** The diseases of DEARGs involved through Disease Ontology analysis. DEARGs, differentially expressed autoimmune-related genes; DO, Disease Ontology.

that the system with the greatest distribution of tissue-specific expressed genes was the hematologic/immune system (30/70, 42.90%). The second organ-specific expressed system was the nervous system, which included 11 genes (11/70, 15.70%), followed by the digestive system (7/70, 10.00%), endocrine system (5/70, 7.10%), and genital system (4/70, 5.70%).

Screening for crucial DEARGs and candidate signatures

First, we integrated five methods (Closeness, Degree, MCC, MNC, and Radiality) in cytoHubba and overlapped the top 30 genes from each method for robustness (Supplementary Figure S2 and Supplementary Table S2). A total of 15 shared genes were identified (Figure 6A). A co-expression network was constructed (Figure 6B), and 20 genes were identified that interacted with the 15 key DEARGs. In the complex PPI network, the interaction of the physical interactions accounted for 38.45%, predicted for 35.98%, co-expression for 20.16%, and colocalization for 3.44%.

Second, we applied three machine learning algorithms (LASSO regression, RF, and SVM-RFE) to screen further the most important signatures. Seven genes were determined by LASSO regression with the optimal values (Figures 7A, B). We selected the top 10 genes ranked by the variable importance in RF (Figure 7C). The error fell to the lowest perigee, and the accuracy reached the peak when the number of features was set to 14 in the SVM-RFE method (Figures 7D, E). We identified six diagnostic genes by overlapping the three algorithms (Figure 7F): *CD38*, *PTPN11*, *NOTCH1*, *TLR7*, *KAT2B*, and *ISG15*. The detailed descriptions of the six diagnostic signatures are listed in Table 2. The correlation among these genes was studied by Pearson's method and presented with a heatmap (Supplementary Figure S3).

TABLE 1 70 Identified tissue/organ-specific expressed genes by BioGPS.

System/Organ	Counts	Frequency (%)	Genes
Haematologic/Immune cells	30	42.90	TNFRSF17, CTSS, FPR2, HLA-DMA, HLA-DRB5, ICAM3, ITGA4, KLRD1, STAT4, MBD4, CARD8, KLRK1, PTPN22, TLR7, CCL4, INPP5D, NLRP12, PYCARD, OAS2, ISG20, KCNJ2, FCHSD2, FUT1, TOP1, CBLN2, XRN2, C3AR1, ZCCHC7, SPATA13, MILR1
Nervous	11	15.70	FGF1, GRIA2, PPP2R2B, TAC1, UCHL1, PPP1R17, SERPINI1, DNAJC6, PDE2A, AAK1, NMNAT2
Digestive	7	10.00	MBL2, SLC22A1, PEMT, CEBPA, SPRY4, APCS, APOA2
Endocrine	5	7.10	CALCA, CA3, FAM47E, SLC26A4, SPINK1
Genital	4	5.70	PGF, PRKAA1, DIAPH2, ARHGAP35
Bonemarrow	3	4.30	BPI, CAMP, DEFA3
Smooth muscle	3	4.30	BDNF, CCL7, CCL11
Adipocyte	2	2.90	COMP, MME
Immune organs	2	2.90	LEF1, NLRP11
Heart	1	1.40	NPPA
Respiratory	1	1.40	LAMP3
Tongue/Skeletal Muscle	1	1.40	TTN

Construction and validation of the biological processes neural network diagnostic model

Based on the above screened six diagnostic signatures, we used the GSE157628 dataset as the training set to construct the back propagation artificial neural network model using the R package "neuralnet". After performing the preprocessing and scaling of this dataset, a neural network model with one hidden neuron layer was established. According to the output results of the neural network model (Figure 8A, Supplementary Table S3, and Supplementary Figure S4), the entire training was performed in 179 steps with an error of 0.021. Among the output results, the predicted weights of each hidden neuron layer were −14.758, −1.778, −5.728, and 13.268 for MMD. The confusion matrix and ROC curves show that the predicted scores (namely neuroMMD) present great diagnostic ability with AUC = 1 in the GSE157628 dataset (Figures 8B, C). The relative expression levels of these six diagnostic signatures are depicted through raincloud plots, indicating that *CD38*, *NOTCH11*, *TLR7*, and *ISG15* were upregulated, and *PTPN11* and *KAT2B* were downregulated (Figure 8D). We also recruited an external GSE141025 dataset that demonstrated the discriminatory performance of neuroMMD scores for MMD. The confusion matrix and ROC curves validated the great diagnostic ability of neuroMMD scores in MMD (Figures 8E, F). Although these genes from the verifying dataset showed the same expression trends as the training dataset, only CD38 and PTPN11 reached statistical significance (Figure 8G). Our RT-qPCR results also verified the same expression trends of CD38 and PTPN11 as the GSE157628 and GSE141025 datasets, and we also found the significant up-regulation of NOTCH1 of MMD patients in our results (Supplementary Figure S5).

Identification of differential immune characteristics between moyamoya disease and controls

Based on xCell analysis results, we identified three immune cells (eosinophils, natural killer T cells, and Th2 cells) with significant infiltration differences between diseased tissues and controls. Among them, eosinophils and NKT were increased in diseased tissues, while Th2 cells were decreased. We also assessed the differences in immune activities and responses between MMD and controls using the ssGSEA algorithm (Figure 9A). Three immune activities (interleukins, interferon, and LCK molecules) showed significant differences between groups (Figure 9B). Seven differential expressed HLA molecules were also discovered (Figure 9C), and all were upregulated. The differential immune characteristics between MMD and controls are depicted through a heatmap (Figure 9D).

The correlation analysis between six hub DEARGs and differential immune characteristics showed that the upregulated genes were positively correlated with increasing immune characteristics and negatively with decreasing immune characteristics. In contrast, the downregulated genes showed inverse effects.

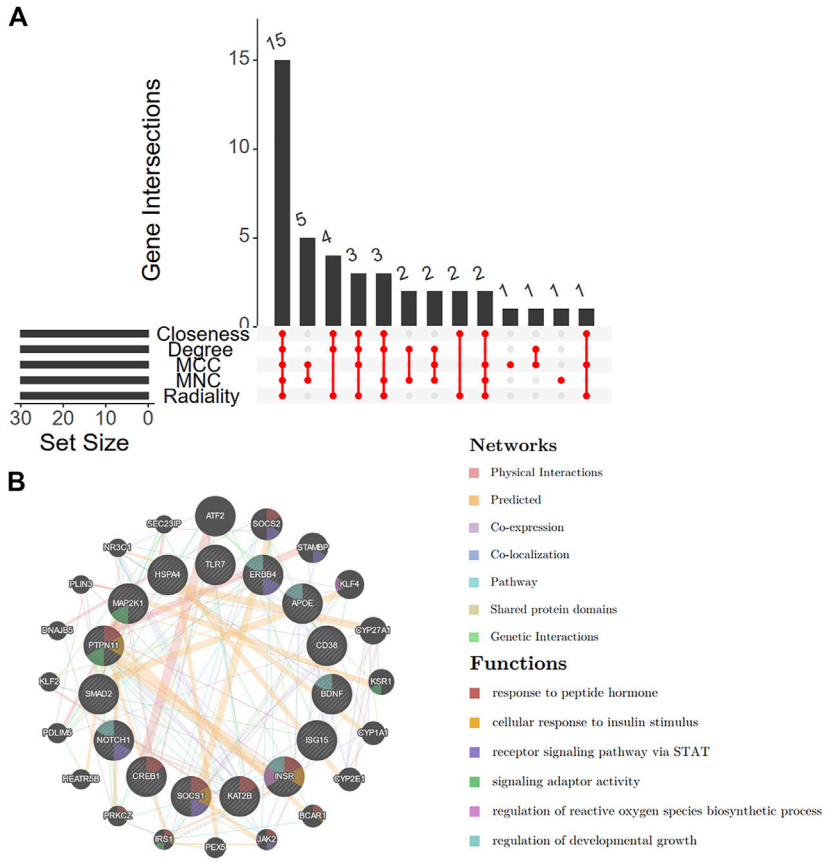


FIGURE 6
Venn diagram and co-expression of key DEARGs. **(A)** A Venn diagram shows that 15 key DEARGs are common genes from five cytoHubba methods. **(B)** 15 key DEARGs and their co-expressed genes analyzed by GENEMANIA. DEARGs, differentially expressed autoimmune-related genes.

Establishment of miRNA-gene and drug-gene regulatory networks

To identify the regulatory and therapeutic mechanisms for MMD, we predicted the miRNA and drugs targeting the six hub DEARGs through the Tarbase and DGIdb databases, respectively. A total of 131 miRNAs were found to potentially regulate the hub ARDEGs (Figure 10A). Genes with the most regulated miRNAs were identified as *KAT2B* (regulated by 56 miRNAs), followed by *PTPN11* (43 miRNAs), *NOTCH1* (35 miRNAs), *ISG15* (29 miRNAs), *CD38* (13 miRNAs), and *TLR7* (5 miRNAs) (Supplementary Table S4). We also discovered that four miRNAs (mir-26a-5p, mir-1343-3p, mir-129-2-3p, and mir-124-3p) interacted with four hub genes.

The DGIdb was applied to predict possible medicines or molecular compounds reacting with the hub ARDEGs. After excluding drugs/compounds for which the interaction types with genes were not clear, 13 potential target drugs/compounds remained for MMD treatment (Figure 10B and Supplementary Table S5). Of these, six drugs interacted with *TLR7*; two targeted *CD38*, *KAT2B*, and *NOTCH1*; one drug targeted *ISG15*. *PTPN11* interacted with no potential drug.

Discussion

As the most common pediatric cerebrovascular disease in Japan, the incidence of frequently recurrent ischemic episodes (transient ischemic attacks or strokes) is 70%–80% in children with MMD (Currie et al., 2011). MMD cannot be effectively treated with pharmacological interventions alone; therefore, surgical procedures for revascularization (direct, indirect, and combined bypass) are required. There is much controversy surrounding the optimal treatment for this disorder (Jang et al., 2017; Deng et al., 2018). The incidence of perioperative complications such as stroke, hyperperfusion syndrome, and acute thrombogenesis is also concerning (Kim et al., 2016). It is of profound significance to understand the pathophysiologic processes of MMD and prevent the occurrence of this disorder. However, the molecular etiology of MMD remains unclear.

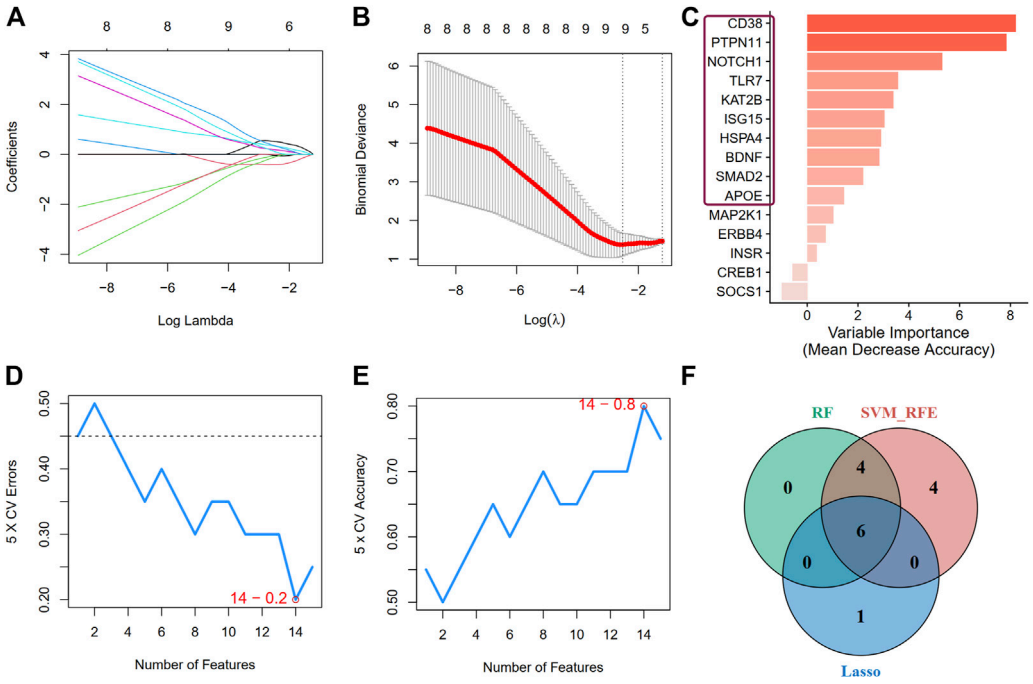


FIGURE 7
Screen for crucial DEARGs based on machine learning algorithms. **(A,B)** Feature selection by LASSO regression model **(A)** The coefficients change of different genes with different lambda **(B)** By verifying the optimal parameter (lambda) in the LASSO model, the partial likelihood deviance (binomial deviance) curve was plotted vs log(lambda). **(C)** The genes ranked by the feature importance based on random forest algorithm. The darker the color, the more important the gene is. **(D)** and **(E)** The error and accuracy of model changed with different number of features in support vector machine-recursive feature elimination method, respectively. **(F)** A Venn diagram demonstrating six diagnostic markers shared by the three algorithms.

TABLE 2 Detail information about the six hub genes identified by machine learning.

Gene	Description	Chromosome	logFC	P.Value	Change
CD38	CD38 molecule	4	1.257	0.029	UP
PTPN11	protein tyrosine phosphatase non-receptor type 11	12	−1.021	0.007	DOWN
NOTCH1	notch receptor 1	9	1.104	0.005	UP
TLR7	toll like receptor 7	X	1.417	0.006	UP
KAT2B	lysine acetyltransferase 2B	3	−1.435	0.046	DOWN
ISG15	ISG15 ubiquitin like modifier	1	1.072	0.023	UP

Previous studies have suggested the role of comorbidities as the link between autoimmune diseases and MMD (Kim et al., 2010; Chen et al., 2016). Therefore, analysis of ARGs may help determine the pathogenesis of MMD.

This study aimed to elucidate the critical processes and ARGs responsible for developing MMD by integrating bioinformatics and machine learning methods. Machine learning and artificial intelligence have become indispensable productivity tools for the 21st century for precision medicine. Machine learning and artificial intelligence differ from traditional biomedical research because they use huge volumes of data to uncover natural laws, which are then applied to medical research. The field of bioinformatics involves the development of computational tools and approaches for acquiring, storing, visualizing, and interpreting medical or biological data. Combining machine learning and bioinformatics will facilitate the generation, analysis, maintenance, and interpretation of information derived from molecular genetics tests. Apart from

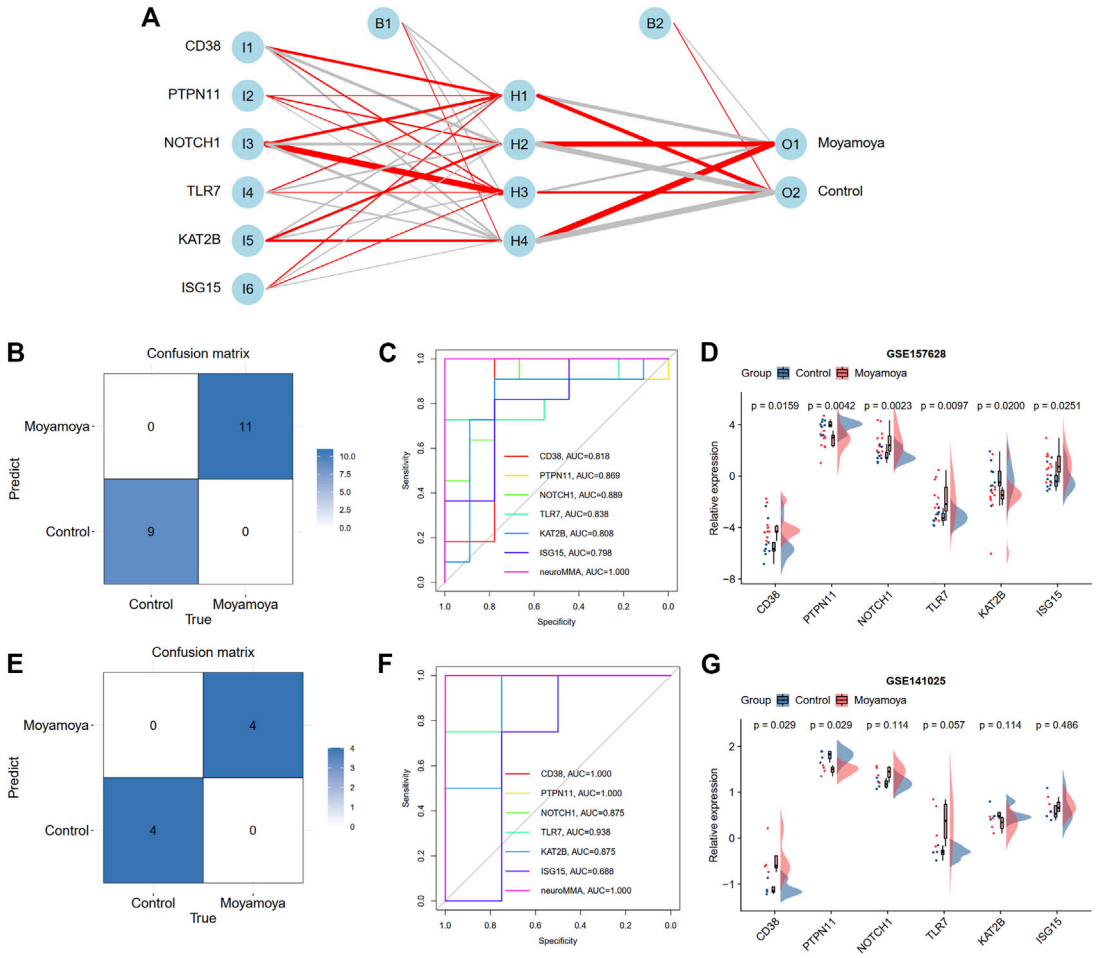


FIGURE 8
Construction and validation a BP neural network diagnostic model. **(A)** Results of neural network visualization. The positive weights are connected with red lines, and the negative weights are connected with gray lines. The thickness of the lines reflects the value of the weights. **(B)** A confusion matrix shows the classification ability of neural network in training dataset. **(C)** ROC curve shows the diagnostic ability of neural network in training dataset. **(D)** The expression levels of six hub DEARGs in training dataset. **(E)** A confusion matrix shows the classification ability of neural network in test dataset. **(F)** ROC curve shows the diagnostic ability of neural network in test dataset. **(G)** The expression levels of six hub DEARGs in test dataset.

the field of oncology research, the integrated approach is widely applied in cardiovascular diseases, such as myocardial infarction (Wu et al., 2022), heart failure (Tian et al., 2020), and aortic valve calcification (Xiong et al., 2022). Through machine learning and bioinformatics technology, our study revealed the differential expression of ARGs and their potential biological functions in MMD for the first time. A predictive model with hub ARGs was constructed by using an artificial neural network. Immune cell infiltration, immune activities, and HLA molecule expression were investigated in MMD. The elucidated correlations between ARGs and immune characteristics may help further explain the interplay of ARGs and the immune microenvironment in MMD.

In our study, a total of 1696 DEGs (750 upregulated genes and 946 downregulated genes) between MMD and controls were screened. In biological function analysis, the pathways of autoimmune diseases (such as autoimmune thyroid disease, rheumatoid arthritis, the intestinal immune network for IgA production, and SLE), cell adhesion molecules (CAMs), and chylomicron remodeling were more enriched in MMD than controls. When overlapped with the ARGs, we further intersected 260 DEARGs involved in bacterial infectious disease, rheumatic disease, and collagen disease in DO analysis. Moyamoya vasculopathy in patients with the underlying causal condition is usually regarded as "MM syndrome". MMD concurrent with Graves' disease (GD) was first reported by Kushima et al. (1991). Over the past two decades, reports of these two concurrent diseases have increased (Tendler et al., 1997; Ni et al., 2014; Chen et al.,

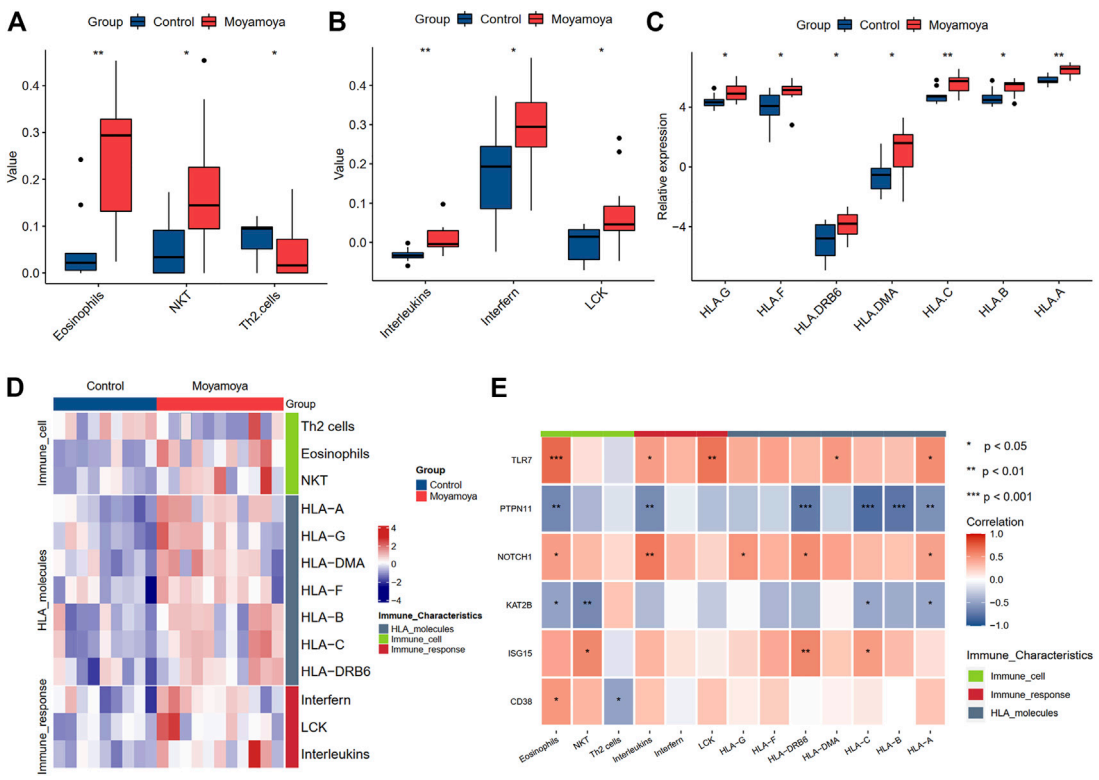


FIGURE 9
The significantly differential immune characteristics between MMD and controls. **(A)** The significantly differential immune cells. **(B)** The significantly differential immune activities. **(C)** The significantly differential expressed HLA molecules. **(D)** A heat map shows the landscapes of immune characteristics between MMD and controls. **(E)** The correlation of immune characteristics and six hub DEARGs. Significance level was denoted by *p-value <0.05, **p-value <0.01, ***p-value <0.001.

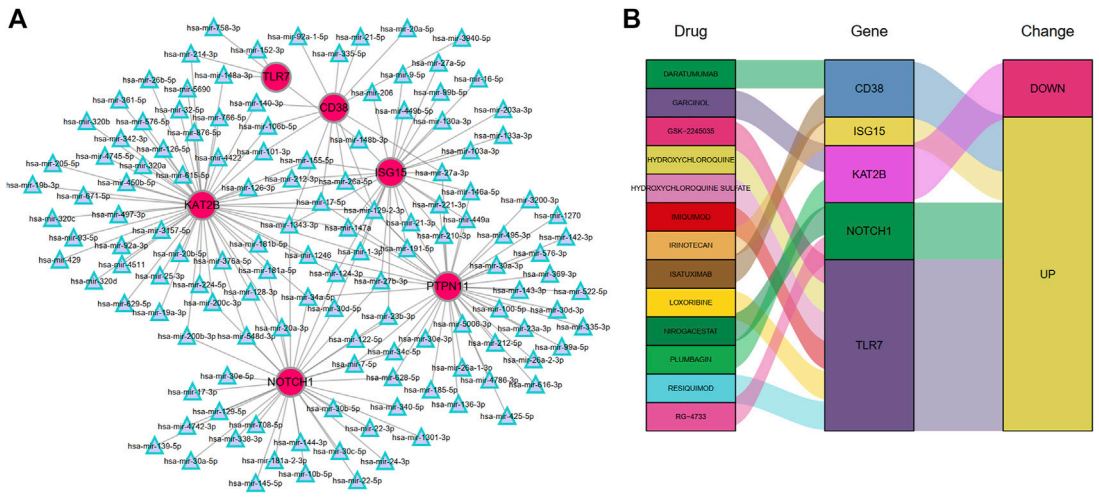


FIGURE 10
The miRNA-Genes regulatory network **(A)** and Drug-Genes interactions **(B)**.

2015). The associations between thyroid function and thyroid autoantibodies with MMD were also discovered (Kim et al., 2010). A case-control study found that compared with control subjects, the thyroid function and thyroid autoantibodies are elevated in pediatric patients with MMD (Li et al., 2011), further supported by another study without age stratification (Lei et al., 2014). Studies about SLE associated with MM syndrome have rarely been conducted and are published mostly in the form of case reports (El Ramahi and Al Rayes, 2000; Jeong et al., 2008). According to a recent review, MMD complicated with SLE mostly occurred in female patients [84.6% (11/13)], and most of these patients developed MMD by the age of 30 years (Tanaka et al., 2020). Among the 13 patients, 10 were from East Asian countries. Complications with rheumatoid arthritis and MMD are rare (Paciaroni et al., 2005). The common molecular characteristic of SLE and rheumatoid arthritis is the change in CAMs (da Rosa Franchi Santos et al., 2020), which was also identified by our enrichment analysis in MMD. CAMs, which are transmembrane proteins that facilitate cell-to-cell or cell-to-extracellular matrix binding, may be categorized into three different types named immunoglobulin supergene family members, selectins, and integrins. CAMs regulate the inflammatory response and endothelial function. Therefore, they may be targeted in cardiovascular disease (Kunutsor et al., 2017), such as atherosclerosis (Ling et al., 2012) and ischemic stroke (Yilmaz and Granger, 2008). DIAPH1 may be a novel MMD risk gene that impairs vascular cell actin remodeling that may cause neointimal expansion and progressive narrowing of the bilateral internal carotid arteries in MMD pathogenesis (Kundishora et al., 2021). Soriano et al. (2002). found significantly elevated levels of soluble CAMs in the cerebrospinal fluid of children with MM syndrome compared with the control group, suggesting the potential roles of CAMs in MMD. These results indicate the crucial association between MMD and autoimmune diseases. By examining the common pathogenesis of these disorders, we can clarify the etiology of MMD. CAMs may act as a bridge that triggers the common pathogenesis processes.

We also performed enrichment analysis of DEARGs in MMD, suggesting that infectious diseases (such as *Staphylococcus aureus* infection, Epstein-Barr virus infection, and *tuberculosis*), cAMP signaling pathway, and PI3K-Akt signaling pathway were involved. The infection hypothesis has occasionally been proposed as one of three mechanisms while investigating the pathogenesis of MMD, apart from autoimmune and HLA abnormality. Infections associated with MMD have been reported in many cases, including bacterial meningitis due to pneumococcus, tuberous infection, and viral infection by the varicella-zoster virus, Epstein-Barr virus, and *Leptospira* infection (Houkin et al., 2012). Czartoski proposed that the inflammation and subsequent post-infectious autoimmune response associated with meningitis can lead to a progressive vasculopathy, which may cause arterial occlusions in MM syndrome after autopsy in a patient with pneumococcal meningitis (Czartoski et al., 2005). Despite suggesting a possible infectious cause in MMD, these results were only based on case studies, and no specific pathogen has been identified. A large-sample study is indispensable to finding the relationship between infections and MMD. The DEARGs we identified as associated with these diseases may provide a molecular-level explanation.

A PPI network was constructed based on the DEARGs to explore relationships among proteins. We found 15 potential genes overlapping the top 30 genes identified by six algorithms (Closeness, Degree, MCC, MNC, and Radiality) in cytoHubba. To further screen out the hub genes, three machine learning methods (LASSO regression, RF, and SVM-RFE) were applied, and six genes (*CD38*, *PTPN11*, *NOTCH1*, *TLR7*, *KAT2B*, and *ISG15*) were selected to construct a predictive model using BP artificial neural network in the GSE157628 dataset. The GSE141025 dataset also verified the predictive performance with great diagnostic ability, which proved the applicability of our model. Type I interferons (IFNs) induce the expression of over 500 genes, collectively referred to as IFN-stimulated genes (ISGs). ISG15 is a ubiquitin-like protein that can conjugate to substrate proteins (ISGylation) in response to microbial infection. This IFN-α/β-inducible ISG15 does not serve as a substrate for ISGylation-based antiviral immunity but for regulating IFN-α/β by USP18 and preventing IFN-α/β-dependent auto-inflammation (Zhang et al., 2015). The antiviral and antineoplastic roles of ISG15 have been extensively studied (Mustachio et al., 2018; Perng and Lenschow, 2018). RNF213 is an interferon-induced mega protein frequently mutated in MMD as a susceptibility gene (Liu et al., 2011). A recent study pointed out that RNF213, an ISG15 interactor, can act as a sensor for ISGylated proteins to counteract infection (Thery et al., 2021). In our immune infiltration analysis, we observed that the activity of IFNs and the expression of ISG15 genes were higher in MMD than in controls. Therefore, the overexpression of ISG15 induced by IFNs may be involved in the pathogenesis of MMD through its interaction with RNF213. This finding may provide a new direction for basic experiments in the future. *PTPN11*, the gene encoding the protein tyrosine phosphatase SHP2, is a ubiquitously expressed non-receptor tyrosine phosphatase that regulates cell survival, proliferation, differentiation, migration, and adhesion. Germline mutations in *PTPN11* cause Noonan syndrome, the clinically related LEOPARD syndrome (LS), and leukemogenesis (Tartaglia et al., 2006; Alfayez et al., 2021). Seki et al. found that the expression of SHP2 was markedly elevated in the thickened aortic intima in rats with balloon-induced injury in an atherosclerosis animal model (Seki et al., 2002). The inhibition of SHP2 can protect against atherosclerosis by inhibiting smooth muscle cell proliferation (Chen et al., 2018). The most prominent pathological change in MMD is the inner elastic lamellar's breakage and smooth muscle cells' destruction and

proliferation in the tunica media (Huang et al., 2017). There are some common pathogenesis links between MMD and atherosclerosis. MMD susceptibility variant RNF213 p. R4810K can increase the risk of recurrent cerebrovascular events, such as ischemic stroke caused by large-artery atherosclerosis (Okazaki et al., 2019; Kim et al., 2021). Considering no direct evidence connecting *CD38*, *PTPN11*, *NOTCH1*, *TLR7*, and *KAT2B* with the pathogenesis of MMD, their roles in atherosclerosis may also provide a new perspective and direction for future research on the molecular targeted therapy of MMD (Salagianni et al., 2012; Briot et al., 2015; Xu et al., 2016; Qi et al., 2021).

Considering the important roles of immune activities in MMD, we also studied the immune characteristics from the perspective of immune cell infiltration, activities of immune responses, and HLA molecule expression. Our results showed that the abundance of eosinophils and natural killer T (NKT) cells is significantly elevated while Th2 cells were decreased in MMD compared to controls. The expression levels of HLA-A, HLA-B, HLA-C, HLA-DMA, HLA-DRB6, HLA-F, and HLA-G were significantly upregulated in MMD. The abnormality of HLA is considered one of the molecular mechanisms leading to the occurrence of MMD. Hong et al. (2009) found that the phenotype frequencies of HLA-DRB1(*) 1302 and DQB1(*) 0609 were significantly increased in familial MMD compared to both controls and non-familial Korean patients with MMD. In a Japanese case-control study on MMD, the HLA-DRB1*04:10 allele was found to be a predisposing genetic factor, and the frequency of autoimmune thyroid diseases was increased in HLA-DRB1*04:10-positive patients with MMD compared with that in HLA-DRB1*04:10-negative patients with MMD (Tashiro et al., 2019). Recent research in Chinese Han population indicated that the genetic polymorphism of HLA-DQA2 and HLA-B was identified as a risk factor for MMD (Wan et al., 2021).

Gene-miRNAs modify the appearance of proteins with the progression of diseases by targeting their main targets. In this study, we also constructed a gene-miRNA regulatory network, and four miRNAs (mir-26a-5p, mir-1343-3p, mir-129-2-3p, and mir-124-3p) were identified because of their interaction at least with four hub DEARGs. Mir-26a-5p can alleviate cardiac hypertrophy and dysfunction (Shi et al., 2021) and protect against myocardial ischemia/reperfusion injury (Xing et al., 2020). Mir-1343-3p plays a significant role in the development of human cancers such as lung cancer (Zhang et al., 2020), colorectal carcinoma (Bhat et al., 2022), and hepatocellular carcinoma (Mou and Sun, 2022). The involvement of mir-1343-3p in cardiovascular diseases is also reported (Sharma et al., 2020). The potential role of mir-129-2-3p in ischemic stroke was identified by suppressing *SYK* gene expression (Huang et al., 2019). Mir-124-3p contributes to the development of different cardiovascular diseases, such as atherosclerosis, myocardial infarction, and ischemic stroke (de Ronde et al., 2017; Badacz et al., 2021). These four miRNAs may be used as interventional targets for examining the mechanisms of ARGs in MMD since they interact with at least four hub DEARGs. Moreover, a total of 13 potential drugs/compounds were predicted for MMD treatment by targeting the hub ARGs in our study.

Further studies are warranted to address some limitations of the present study. First, although the diagnostic model constructed by an artificial neural network performed well in the training and testing datasets, the sample size was very small, especially for the validation dataset, where only four samples were available in each group. Therefore, studies with larger sample sizes are essential. Second, in the GSE157628 dataset, six control samples were collected from the MCA of patients with ICA aneurysms. Considering the different hemodynamic and genetic effects, the normal artery of patients with aneurysms may differ from normal vessels at the transcriptional level. However, the collection of normal vessels from healthy control, in essence, is against medical ethics. Setting the normal artery from patients with aneurysms as the control group in studying MMD is acceptable (Kanamori et al., 2021). Third, the results were based on bioinformatics and conducted RT-qPCR, but *in vitro* and *in vivo* experiments should be conducted to verify the results.

Conclusion

In our analysis, a total of 260 DEARGs were identified in MMD, which were involved in autoimmune-related diseases and immune responses. Six ARGs (*CD38*, *PTPN11*, *NOTCH1*, *TLR7*, *KAT2B*, and *ISG15*) were selected by three machine learning methods (LASSO regression, RF, and SVM-RFE). They were finally used to construct a predictive model using BP artificial neural network that could be used to identify patients with MMD. Immune infiltration analysis showed that immune activities and HLA expression levels in MMD were enhanced. Finally, a gene-miRNA network was prepared, and pharmacological agents targeting hub genes were predicted as potentially effective in treating MMD.

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

Conceptualization, SL and YH; methodology, SL; software, SL; validation, YH; formal analysis, SL; investigation, SL; writing—original draft preparation, SL and YH;

writing—review and editing, DT and QZ; visualization, YH and JL; supervision, LW.

Funding

This research was supported by National Natural Science Foundation of China (No. 81902551) and Hunan Provincial Natural Science Foundation of China (No. 2021JJ31072).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.991425/full#supplementary-material

References

Alfayez, M., Issa, G. C., Patel, K. P., Wang, F., Wang, X., Short, N. J., et al. (2021). The Clinical impact of PTPN11 mutations in adults with acute myeloid leukemia. *Leukemia* 35, 691–700. doi:10.1038/s41375-020-0920-z

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18, 220. doi:10.1186/s13059-017-1349-1

Badacz, R., Kleczyński, P., Legutko, J., Żmudka, K., Gacoń, J., Przewłocki, T., et al. (2021). Expression of miR-1-3p, miR-16-5p and miR-122-5p as possible risk factors of secondary cardiovascular events. *Biomedicines* 9, 1055. doi:10.3390/biomedicines9081055

Bhat, I. P., Rather, T. B., Bhat, G. A., Maqbool, I., Akhtar, K., Rashid, G., et al. (2022). TEAD4 nuclear localization and regulation by miR-4269 and miR-1343-3p in colorectal carcinoma. *Pathol. Res. Pract.* 231, 153791. doi:10.1016/j.prp.2022.153791

Briot, A., Civelek, M., Seki, A., Hoi, K., Mack, J. J., Lee, S. D., et al. (2015). Endothelial NOTCH1 is suppressed by circulating lipids and antagonizes inflammation during atherosclerosis. *J. Exp. Med.* 212, 2147–2163. doi:10.1084/jem.20150603

Chen, J. B., Lei, D., He, M., Sun, H., Liu, Y., Zhang, H., et al. (2015). Clinical features and disease progression in moyamoya disease patients with Graves disease. *J. Neurosurg.* 123, 848–855. doi:10.3171/2014.10.Jns141140

Chen, J. B., Liu, Y., Zhou, L. X., Sun, H., He, M., and You, C. (2016). Increased prevalence of autoimmune disease in patients with unilateral compared with bilateral moyamoya disease. *J. Neurosurg.* 124, 1215–1220. doi:10.3171/2015.4.Jns142936

Chen, J., Cao, Z., and Guan, J. (2018). SHP2 inhibitor PHPS1 protects against atherosclerosis by inhibiting smooth muscle cell proliferation. *BMC Cardiovasc. Disord.* 18, 72. doi:10.1186/s12872-018-0816-2

Chen, B., Tao, W., Li, S., Zeng, M., Zhang, L., Huang, Z., et al. (2022). Medial gap: a structural factor at the arterial bifurcation aggravating hemodynamic insult. *J. Neuropathol. Exp. Neurol.* 81, 282–290. doi:10.1093/jnen/nlac017

Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8, S11. doi:10.1186/1752-0509-8-s4-s11

Currie, S., Raghavan, A., Batty, R., Connolly, D. J., and Griffiths, P. D. (2011). Childhood moyamoya disease and moyamoya syndrome: a pictorial review. *Pediatr. Neurol.* 44, 401–413. doi:10.1016/j.pediatrneurol.2011.02.007

Czartoski, T., Hallam, D., Lacy, J. M., Chun, M. R., and Becker, K. (2005). Postinfectious vasculopathy with evolution to moyamoya syndrome. *J. Neurol. Neurosurg. Psychiatry* 76, 256–259. doi:10.1136/jnnp.2004.041046

da Rosa Franchi Santos, L. F., Costa, N. T., Maes, M., Simão, A. N. C., and Dichi, I. (2020). Influence of treatments on cell adhesion molecules in patients with systemic lupus erythematosus and rheumatoid arthritis: a review. *Inflammopharmacology* 28, 363–384. doi:10.1007/s10787-019-00674-6

de Ronde, M. W. J., Kok, M. G. M., Moerland, P. D., Van den Bossche, J., Neele, A. E., Halliani, A., et al. (2017). High miR-124-3p expression identifies smoking individuals susceptible to atherosclerosis. *Atherosclerosis* 263, 377–384. doi:10.1016/j.atherosclerosis.2017.03.045

Deng, X., Gao, F., Zhang, D., Zhang, Y., Wang, R., Wang, S., et al. (2018). Effects of different surgical modalities on the clinical outcome of patients with moyamoya disease: a prospective cohort study. *J. Neurosurg.* 128, 1327–1337. doi:10.3171/2016.12.Jns162626

El Ramahi, K. M., and Al Rayes, H. M. (2000). Systemic lupus erythematosus associated with moyamoya syndrome. *Lupus* 9, 632–636. doi:10.1191/096120300678828686

Hong, S. H., Wang, K. C., Kim, S. K., Cho, B. K., and Park, M. H. (2009). Association of HLA-DR and -DQ genes with familial moyamoya disease in Koreans. *J. Korean Neurosurg. Soc.* 46, 558–563. doi:10.3340/jkns.2009.46.6.558

Houkin, K., Ito, M., Sugiyama, T., Shichinohe, H., Nakayama, N., Kazumata, K., et al. (2012). Review of past research and current concepts on the etiology of moyamoya disease. *Neurol. Med. Chir.* 52, 267–277. doi:10.2176/nmc.52.267

Huang, S., Guo, Z. N., Shi, M., Yang, Y., and Rao, M. (2017). Etiology and pathogenesis of Moyamoya Disease: An update on disease prevalence. *Int. J. Stroke* 12, 246–253. doi:10.1177/1747493017694393

Huang, S., Lv, Z., Wen, Y., Wei, Y., Zhou, L., Ke, Y., et al. (2019). miR-129-2-3p directly targets SYK gene and associates with the risk of ischaemic stroke in a Chinese population. *J. Cell. Mol. Med.* 23, 167–176. doi:10.1111/jcmm.13901

Jang, D. K., Lee, K. S., Rha, H. K., Huh, P. W., Yang, J. H., Park, I. S., et al. (2017). Bypass surgery versus medical treatment for symptomatic moyamoya disease in adults. *J. Neurosurg.* 127, 492–502. doi:10.3171/2016.8.Jns152875

Jeong, H. C., Kim, Y. J., Yoon, W., Joo, S. P., Lee, S. S., and Park, Y. W. (2008). Moyamoya syndrome associated with systemic lupus erythematosus. *Lupus* 17, 679–682. doi:10.1177/0961203307087375

Jin, F., and Duan, C. (2022). Identification of immune-infiltrated hub genes as potential biomarkers of Moyamoya disease by bioinformatics analysis. *Orphanet J. Rare Dis.* 17, 80. doi:10.1186/s13023-022-02238-4

Kanamori, F., Yokoyama, K., Ota, A., Yoshikawa, K., Karnan, S., Maruwaka, M., et al. (2021). Transcriptome-wide analysis of intracranial artery in patients with moyamoya disease showing upregulation of immune response, and downregulation of oxidative phosphorylation and DNA repair. *Neurosurg. Focus* 51, E3. doi:10.3171/2021.6.Focus20870

Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., et al. (2018). DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.* 46, D239–d45. doi:10.1093/nar/gkx1141

Kim, S. J., Heo, K. G., Shin, H. Y., Bang, O. Y., Kim, G. M., Chung, C. S., et al. (2010). Association of thyroid autoantibodies with moyamoya-type cerebrovascular disease: a prospective study. *Stroke* 41, 173–176. doi:10.1161/strokeaha.109.562264

Kim, T., Oh, C. W., Bang, J. S., Kim, J. E., and Cho, W. S. (2016). Moyamoya disease: Treatment and outcomes. *J. Stroke* 18, 21–30. doi:10.5853/jos.2015.01739

Kim, H. J., Choi, E. H., Chung, J. W., Kim, J. H., Kim, Y. S., Seo, W. K., et al. (2021). Role of the RNF213 variant in vascular outcomes in patients with intracranial atherosclerosis. *J. Am. Heart Assoc.* 10, e017660. doi:10.1161/jaha.120.017660

Kundishora, A. J., Peters, S. T., Pinard, A., Duran, D., Panchagnula, S., Barak, T., et al. (2021). DIAPH1 variants in non-east Asian patients with sporadic moyamoya disease. *JAMA Neurol.* 78, 993–1003. doi:10.1001/jamaneurol.2021.1681

Kunutsor, S. K., Bakker, S. J. L., and Dullaart, R. P. F. (2017). Soluble vascular cell adhesion molecules may be protective of future cardiovascular disease risk: Findings from the PREVEND prospective cohort study. *J. Atheroscler. Thromb.* 24, 804–818. doi:10.5551/jat.38836

Kuroda, S., and Houkin, K. (2008). Moyamoya disease: current concepts and future perspectives. *Lancet. Neurol.* 7, 1056–1066. doi:10.1016/s1474-4422(08)70240-0

Kushima, K., Satoh, Y., Ban, Y., TaniyaMaM.Ito, K., and Sugita, K. (1991). Graves' thyrotoxicosis and Moyamoya disease. *Can. J. Neurol. Sci.* 18, 140–142. doi:10.1017/s0317167100031607

Lei, C., Wu, B., Ma, Z., Zhang, S., and Liu, M. (2014). Association of moyamoya disease with thyroid autoantibodies and thyroid function: a case-control study and meta-analysis. *Eur. J. Neurol.* 21, 996–1001. doi:10.1111/ene.12425

Li, H., Zhang, Z. S., Dong, Z. N., Ma, M. J., Yang, W. Z., Han, C., et al. (2011). Increased thyroid function and elevated thyroid autoantibodies in pediatric patients with moyamoya disease: a case-control study. *Stroke* 42, 1138–1139. doi:10.1161/strokeaha.110.608471

Li, S., Zhang, Q., Chen, Z., Huang, Z., and Chen, F. (2022). Novel insight into ferroptosis-related genes, molecular subtypes, and immune characteristics in intracranial aneurysms. *Inflamm. Res.* 71, 1347–1364. doi:10.1007/s00011-022-01633-8

Li, S., Zhang, Q., Weng, L., Han, Y., and Li, J. (2022). Novel insight into m6A regulator-mediated methylation modification patterns and immune characteristics in intracranial aneurysm. *Front. Aging Neurosci.* 14, 973258. doi:10.3389/fnagi.2022.973258

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Ling, S., Nheu, L., and Komesaroff, P. A. (2012). Cell adhesion molecules as pharmaceutical target in atherosclerosis. *Mini Rev. Med. Chem.* 12, 175–183. doi:10.2174/138955712798995057

Liu, W., Morito, D., Takashima, S., Mineharu, Y., Kobayashi, H., Hitomi, T., et al. (2011). Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. *PloS one* 6, e22542. doi:10.1371/journal.pone.0022542

Lu, G., Hao, X., Chen, W. H., and Mu, S. (2018). GAAD: A gene and autoimmiune disease association database. *Genomics Proteomics Bioinforma.* 16, 252–261. doi:10.1016/j.gpb.2018.05.001

Mou, Y., and Sun, Q. (2022). The long non-coding RNA ASMTL-AS1 promotes hepatocellular carcinoma progression by sponging miR-1343-3p that suppresses LAMC1 (laminin subunit gamma 1). *Bioengineered* 13, 746–758. doi:10.1080/21655979.2021.2012628

Mustachio, L. M., Lu, Y., Kawakami, M., Roszik, J., Freemantle, S. J., Liu, X., et al. (2018). Evidence for the ISG15-specific deubiquitinase USP18 as an antineoplastic target. *Cancer Res.* 78, 587–592. doi:10.1158/0008-5472.Can-17-1752

Ni, J., Zhou, L. X., Wei, Y. P., Li, M. L., Xu, W. H., Gao, S., et al. (2014). Moyamoya syndrome associated with Graves' disease: a case series study. *Ann. Transl. Med.* 2, 77. doi:10.3978/j.issn.2305-5839.2014.08.03

Okazaki, S., Morimoto, T., Kamatani, Y., Kamimura, T., Kobayashi, H., Harada, K., et al. (2019). Moyamoya disease susceptibility variant RNF213 p.R4810K increases the risk of ischemic stroke attributable to large-artery atherosclerosis. *Circulation* 139, 295–298. doi:10.1161/circulationaha.118.038439

Paciaroni, M., Micheli, S., Caso, V., Venti, M., Alberti, A., Milia, P., et al. (2005). Angiographic findings of moyamoya vessels in a patient with rheumatoid arthritis. *Cerebrovasc. Dis.* 20, 415–416. doi:10.1159/000088666

Perng, Y. C., and Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* 16, 423–439. doi:10.1038/s41579-018-0020-5

Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855. doi:10.1093/nar/gkz1021

Qi, X., Wang, H., Xia, L., Lin, R., Li, T., Guan, C., et al. (2021). miR-30b-5p releases HMGB1 via UBE2D2/KAT2B/HMGB1 pathway to promote pro-inflammatory polarization and recruitment of macrophages. *Atherosclerosis* 324, 38–45. doi:10.1016/j.atherosclerosis.2021.02.016

Salagianni, M., Galani, I. E., Lundberg, A. M., Davos, C. H., Varela, A., Gavriil, A., et al. (2012). Toll-like receptor 7 protects from atherosclerosis by constraining "inflammatory" macrophage activation. *Circulation* 126, 952–962. doi:10.1161/circulationaha.111.067678

Scott, R. M., and Smith, E. R. (2009). Moyamoya disease and moyamoya syndrome. *N. Engl. J. Med.* 360, 1226–1237. doi:10.1056/NEJMra0804622

Seki, N., Hashimoto, N., Suzuki, Y., Mori, S., Amano, K., and Saito, Y. (2002). Role of SRC homology 2-containing tyrosine phosphatase 2 on proliferation of rat smooth muscle cells. *Arterioscler. Thromb. Vasc. Biol.* 22, 1081–1085. doi:10.1161/01.atv.0000022878.37277.ec

Sharma, A. R., Vohra, M., Shukla, V., Guddattu, V., Razak Uk, A., Shetty, R., et al. (2020). Coding SNPs in hsa-miR-1343-3p and hsa-miR-6783-3p target sites of CYP2C19 modulates clopidogrel response in individuals with cardiovascular diseases. *Life Sci.* 245, 117364. doi:10.1016/j.lfs.2020.117364

Shi, H., Li, H., Zhang, F., Xue, H., Zhang, Y., and Han, Q. (2021). MiR-26a-5p alleviates cardiac hypertrophy and dysfunction via targeting ADAM17. *Cell Biol. Int.* 45, 2357–2367. doi:10.1002/cbin.11685

Soriano, S. G., Cowan, D. B., Proctor, M. R., and Scott, R. M. (2002). Levels of soluble adhesion molecules are elevated in the cerebrospinal fluid of children with moyamoya syndrome. *Neurosurgery* 50, 544–549. doi:10.1097/00006123-200203000-00022

Suzuki, J., and Takaku, A. (1969). Cerebrovascular "moyamoya" disease. Disease showing abnormal net-like vessels in base of brain. *Arch. Neurol.* 20, 288–299. doi:10.1001/archneur.1969.00480090076012

Tanaka, R., Shimojima, Y., Ueno, K. I., Takasone, K., Ichikawa, T., Kishida, D., et al. (2020). Moyamoya syndrome related to systemic lupus erythematosus developing during pregnancy: a case-based review. *Clin. Rheumatol.* 39, 3861–3867. doi:10.1007/s10067-020-05246-6

Tartaglia, M., Martinelli, S., Stella, L., Bocchinfuso, G., Flex, E., Cordeddu, V., et al. (2006). Diversity and functional consequences of germline and somatic PTPN11 mutations in human disease. *Am. J. Hum. Genet.* 78, 279–290. doi:10.1086/499925

Tashiro, R., Niizuma, K., Khor, S. S., Tokunaga, K., Fujimura, M., Sakata, H., et al. (2019). Identification of HLA-DRB1*04:10 allele as risk allele for Japanese moyamoya disease and its association with autoimmune thyroid disease: A case-control study. *PloS one* 14, e0220858. doi:10.1371/journal.pone.0220858

Tendler, B. E., Shoukri, K., Malchoff, C., MacGillivray, D., DuckRow, R., Talmadge, T., et al. (1997). Concurrence of Graves' disease and dysplastic cerebral blood vessels of the moyamoya variety. *Thyroid* 7, 625–629. doi:10.1089/thy.1997.7.625

Thery, F., Martina, L., Asselman, C., Zhang, Y., Vessely, M., Repo, H., et al. (2021). Ring finger protein 213 assembles into a sensor for ISGylated proteins with antimicrobial activity. *Nat. Commun.* 12, 5772. doi:10.1038/s41467-021-26061-w

Tian, Y., Yang, J., Lan, M., and Zou, T. (2020). Construction and analysis of a joint diagnosis model of random forest and artificial neural network for heart failure. *Aging* 12, 26221–26235. doi:10.18632/aging.202405

Wan, J., Ling, W., Zhengshan, Z., Xianbo, Z., Lian, D., and Kai, W. (2021). Association of HLA-DQA2 and HLA-B with moyamoya disease in the Chinese han population. *Neurol. Genet.* 7, e592. doi:10.1212/nxg.0000000000000592

Wang, H., Zhu, H., Zhu, W., Xu, Y., Wang, N., Han, B., et al. (2020). Bioinformatic analysis identifies potential key genes in the pathogenesis of turner syndrome. *Front. Endocrinol.* 11, 104. doi:10.3389/fendo.2020.00104

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220. doi:10.1093/nar/gkq537

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., et al. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* 10, R130. doi:10.1186/gb-2009-10-11-r130

Wu, Y., Chen, H., Li, L., Zhang, L., Dai, K., Wen, T., et al. (2022). Construction of novel gene signature-based predictive model for the diagnosis of acute myocardial infarction by combining random forest with artificial neural network. *Front. Cardiovasc. Med.* 9, 876543. doi:10.3389/fcvm.2022.876543

Xing, X., Guo, S., Zhang, G., Liu, Y., Wang, X., Lu, Q., et al. (2020). miR-26a-5p protects against myocardial ischemia/reperfusion injury by regulating the PTEN/PI3K/AKT signaling pathway. *Braz. J. Med. Biol. Res.* 53, e9106. doi:10.1590/1414-431x20199106

Xiong, T., Han, S., Pu, L., Zhang, T. C., Zhan, X., Fu, T., et al. (2022). Bioinformatics and machine learning methods to identify FN1 as a novel biomarker of aortic valve calcification. *Front. Cardiovasc. Med.* 9, 832591. doi:10.3389/fcvm.2022.832591

Xu, X., Yuan, X., Li, N., Dewey, W. L., Li, P. L., and Zhang, F. (2016). Lysosomal cholesterol accumulation in macrophages leading to coronary atherosclerosis in CD38(-/-) mice. *J. Cell. Mol. Med.* 20, 1001–1013. doi:10.1111/jcmm.12788

Yilmaz, G., and Granger, D. N. (2008). Cell adhesion molecules and ischemic stroke. *Neurol. Res.* 30, 783–793. doi:10.1179/174313208x341085

Yu, G., Wang, L. G., Yan, G. R., and He, Q. Y. (2015). DOSE: an R/bioconductor package for disease ontology semantic and enrichment analysis. *Bioinforma. Oxf. Engl.* 31, 608–609. doi:10.1093/bioinformatics/btu684

Zhang, X., Bogunovic, D., Payelle-Brogard, B., Francois-Newton, V., Speer, S. D., Yuan, C., et al. (2015). Human intracellular ISG15 prevents interferon-α/β over-amplification and auto-inflammation. *Nature* 517, 89–93. doi:10.1038/nature13801

Zhang, X., Du, L., Han, J., Li, X., Wang, H., Zheng, G., et al. (2020). Novel long non-coding RNA LINC02323 promotes epithelial-mesenchymal transition and metastasis via sponging miR-1343-3p in lung adenocarcinoma. *Thorac. Cancer* 11, 2506–2516. doi:10.1111/1759-7714.13562

Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N., and Xia, J. (2019). NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* 47, W234–W241. doi:10.1093/nar/gkz240

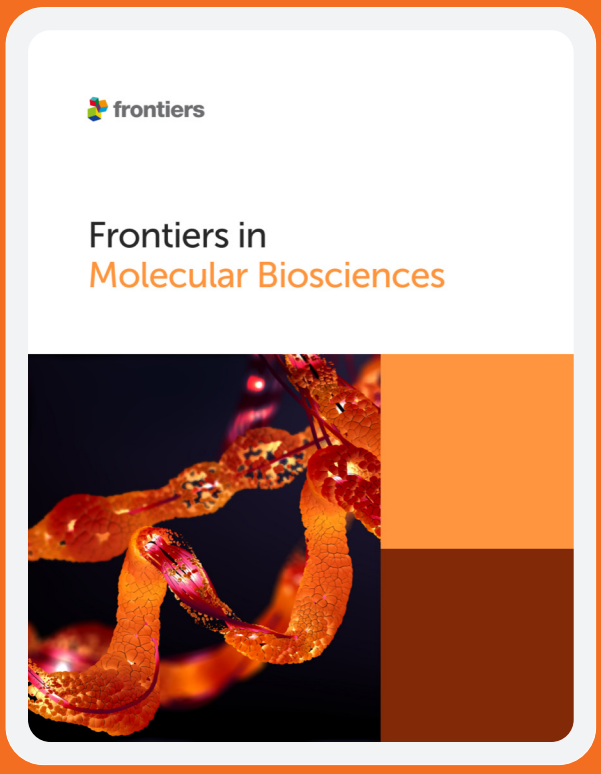
Frontiers in
Molecular Biosciences

Explores biological processes in living organisms on a molecular scale

Focuses on the molecular mechanisms underpinning and regulating biological processes in organisms across all branches of life.

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