ADVANCED COMPUTATIONAL INTELLIGENCE METHODS FOR PROCESSING BRAIN IMAGING DATA

EDITED BY: Kaijian Xia, Yizhang Jiang, Yu-Dong Zhang,
Mohammad Khosravi and Yuanpeng Zhang

PUBLISHED IN: Frontiers in Neuroscience and Frontiers in Aging Neuroscience



frontiers Research Topics

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 978-2-83250-462-8
DOI 10.3389/978-2-83250-462-8

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

ADVANCED COMPUTATIONAL INTELLIGENCE METHODS FOR PROCESSING BRAIN IMAGING DATA

Topic Editors:
Kaijian Xia, Changshu No.1 People's Hospital, China
Yizhang Jiang, Jiangnan University, China
Yu-Dong Zhang, University of Leicester, United Kingdom
Mohammad Khosravi, Persian Gulf University, Iran
Yuanpeng Zhang, Nantong University, China

Citation: Xia, K., Jiang, Y., Zhang, Y.-D., Khosravi, M., Zhang, Y., eds. (2022). Advanced Computational Intelligence Methods for Processing Brain Imaging Data. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-462-8

Table of Contents

08 ***LatLRR-FCNs: Latent Low-Rank Representation With Fully Convolutional Networks for Medical Image Fusion***
Zhengyuan Xu, Wentao Xiang, Songsheng Zhu, Rui Zeng, Cesar Marquez-Chin, Zhen Chen, Xianqing Chen, Bin Liu and Jianqing Li

25 ***Edge-Aware Pyramidal Deformable Network for Unsupervised Registration of Brain MR Images***
Yiqin Cao, Zhenyu Zhu, Yi Rao, Chenchen Qin, Di Lin, Qi Dou, Dong Ni and Yi Wang

39 ***An Inter- and Intra-Subject Transfer Calibration Scheme for Improving Feedback Performance of Sensorimotor Rhythm-Based BCI Rehabilitation***
Lei Cao, Shugeng Chen, Jie Jia, Chunjiang Fan, Haoran Wang and Zhixiong Xu

48 ***GaborNet Visual Encoding: A Lightweight Region-Based Visual Encoding Model With Good Expressiveness and Biological Interpretability***
Yibo Cui, Kai Qiao, Chi Zhang, Linyuan Wang, Bin Yan and Li Tong

60 ***Artificial Intelligence Segmented Dynamic Video Images for Continuity Analysis in the Detection of Severe Cardiovascular Disease***
Xi Zhu, Wei Xia, Zhuqing Bao, Yaohui Zhong, Yu Fang, Fei Yang, Xiaohua Gu, Jing Ye and Wennuo Huang

75 ***Research on the Modality Transfer Method of Brain Imaging Based on Generative Adversarial Network***
Dapeng Cheng, Nuan Qiu, Feng Zhao, Yanyan Mao and Chengnuo Li

87 ***Quantification of Cognitive Function in Alzheimer's Disease Based on Deep Learning***
Yanxian He, Jun Wu, Li Zhou, Yi Chen, Fang Li and Hongjin Qian

105 ***A Novel Unit-Based Personalized Fingerprint Feature Selection Strategy for Dynamic Functional Connectivity Networks***
Feng Zhao, Zhiyuan Chen, Islem Rekik, Peiqiang Liu, Ning Mao, Seong-Whan Lee and Dinggang Shen

117 ***Motor Imagery EEG Signal Recognition Using Deep Convolution Neural Network***
Xiongliang Xiao and Yuee Fang

128 ***A Novel Brain MRI Image Segmentation Method Using an Improved Multi-View Fuzzy c-Means Clustering Algorithm***
Lei Hua, Yi Gu, Xiaoqing Gu, Jing Xue and Tongguang Ni

140 ***Spiking Neural Network for Augmenting Electroencephalographic Data for Brain Computer Interfaces***
Sai Kalyan Ranga Singanamalla and Chin-Teng Lin

152 ***Bidirectional Mapping of Brain MRI and PET With 3D Reversible GAN for the Diagnosis of Alzheimer's Disease***
Wanyun Lin, Weiming Lin, Gang Chen, Hejun Zhang, Qinquan Gao, Yechong Huang, Tong Tong, Min Du and the Alzheimer's Disease Neuroimaging Initiative

165 ***Voxel-Wise Feature Selection Method for CNN Binary Classification of Neuroimaging Data***
Domenico Messina, Pasquale Borrelli, Paolo Russo, Marco Salvatore and Marco Aiello

177 ***A Novel Smart Motor Imagery Intention Human-Computer Interaction Model Using Extreme Learning Machine and EEG Signals***
Yi Gu and Lei Hua

186 ***Advanced Analysis of Diffusion Tensor Imaging Along With Machine Learning Provides New Sensitive Measures of Tissue Pathology and Intra-Lesion Activity in Multiple Sclerosis***
Olayinka Oladosu, Wei-Qiao Liu, Bruce G. Pike, Marcus Koch, Luanne M. Metz and Yunyan Zhang

199 ***Multi-Source Co-adaptation for EEG-Based Emotion Recognition by Mining Correlation Information***
Jianwen Tao and Yufang Dan

218 ***A Transfer Model Based on Supervised Multi-Layer Dictionary Learning for Brain Tumor MRI Image Recognition***
Yi Gu and Kang Li

228 ***Brain Tumor MR Image Classification Using Convolutional Dictionary Learning With Local Constraint***
Xiaoqing Gu, Zongxuan Shen, Jing Xue, Yiqing Fan and Tongguang Ni

240 ***Multimodal Medical Supervised Image Fusion Method by CNN***
Yi Li, Junli Zhao, Zhihan Lv and Zhenkuan Pan

250 ***Characterizing Brain Tumor Regions Using Texture Analysis in Magnetic Resonance Imaging***
Yun Yu, Xi Wu, Jiu Chen, Gong Cheng, Xin Zhang, Cheng Wan, Jie Hu, Shumei Miao, Yuechuchu Yin, Zhongmin Wang, Tao Shan, Shenqi Jing, Wenming Wang, Jianjun Guo, Xinhua Hu and Yun Liu

257 ***Disrupted Asymmetry of Inter- and Intra-Hemispheric Functional Connectivity at Rest in Medication-Free Obsessive-Compulsive Disorder***
Cuicui Jia, Yangpan Ou, Yunhui Chen, Jidong Ma, Chuang Zhan, Dan Lv, Ru Yang, Tinghuizi Shang, Lei Sun, Yuhua Wang, Guangfeng Zhang, Zhenghai Sun, Wei Wang, Xiaoping Wang, Wenbin Guo and Ping Li

265 ***DeepNavNet: Automated Landmark Localization for Neuronavigation***
Christine A. Edwards, Abhinav Goyal, Aaron E. Rusheen, Abbas Z. Kouzani and Kendall H. Lee

277 ***Possibilistic Clustering-Promoting Semi-Supervised Learning for EEG-Based Emotion Recognition***
Yufang Dan, Jianwen Tao, Jianjing Fu and Di Zhou

290 ***Monitoring and Prognostic Analysis of Severe Cerebrovascular Diseases Based on Multi-Scale Dynamic Brain Imaging***
Suting Zhong, Kai Sun, Xiaobing Zuo and Aihong Chen

307 ***Semi-Supervised Support Vector Machine for Digital Twins Based Brain Image Fusion***
Zhibo Wan, Youqiang Dong, Zengchen Yu, Haibin Lv and Zhihan Lv

319 ***Fuzzy System Based Medical Image Processing for Brain Disease Prediction***
Mandong Hu, Yi Zhong, Shuxuan Xie, Haibin Lv and Zhihan Lv

332 ***A Review on Signal Processing Approaches to Reduce Calibration Time in EEG-Based Brain–Computer Interface***
Xin Huang, Yilu Xu, Jing Hua, Wenlong Yi, Hua Yin, Ronghua Hu and Shiyi Wang

351 ***Identification of Epileptic EEG Signals Through TSK Transfer Learning Fuzzy System***
Zhaoliang Zheng, Xuan Dong, Jian Yao, Leyuan Zhou, Yang Ding and Aiguo Chen

363 ***Interpretable Recognition for Dementia Using Brain Images***
Xinjian Song, Feng Gu, Xiude Wang, Songhua Ma and Li Wang

372 ***Improvement Using Planomics Features on Prediction and Classification of Patient-Specific Quality Assurance Using Head and Neck Volumetric Modulated Arc Therapy Plan***
Bing Li, Junying Chen, Wei Guo, Ronghu Mao, Xiaoli Zheng, Xiuyan Cheng, Tiantian Cui, Zhaoyang Lou, Ting Wang, Dingjie Li, Hongyan Tao, Hongchang Lei and Hong Ge

381 ***Evolutionary Multitasking-Based Multiobjective Optimization Algorithm for Channel Selection in Hybrid Brain Computer Interfacing Systems***
Tianyu Liu, Zhixiong Xu, Lei Cao and Guowei Tan

392 ***Automatic Seizure Classification Based on Domain-Invariant Deep Representation of EEG***
Xincheng Cao, Bin Yao, Binqiang Chen, Weifang Sun and Guowei Tan

400 ***A Generalizable and Discriminative Learning Method for Deep EEG-Based Motor Imagery Classification***
Xiuyu Huang, Nan Zhou and Kup-Sze Choi

410 ***Kernelized k-Local Hyperplane Distance Nearest-Neighbor Model for Predicting Cerebrovascular Disease in Patients With End-Stage Renal Disease***
Xiaobin Liu, Xiran Zhang, Yi Zhang, Yijie Ding, Weiwei Shan, Yiqing Huang, Liang Wang and Xiaoyi Guo

416 ***A Novel Active Rehabilitation Model for Stroke Patients Using Electroencephalography Signals and Deep Learning Technology***
Ming Gao and Jie Mao

426 ***Suppressing of Power Line Artifact From Electroencephalogram Measurements Using Sparsity in Frequency Domain***
Jin-Lin Tan, Zhi-Feng Liang, Rui Zhang, You-Qiang Dong, Guang-Hui Li, Min Zhang, Hai Wang and Na Xu

435 ***Selective Cross-Subject Transfer Learning Based on Riemannian Tangent Space for Motor Imagery Brain-Computer Interface***
Yilu Xu, Xin Huang and Quan Lan

449 ***Voxel-Wise Adversarial FiboNet for 3D Cerebrovascular Segmentation on Magnetic Resonance Angiography Images***
Bin Guo, Fugen Zhou, Bo Liu and Xiangzhi Bai

463 ***Electroencephalogram-Based Motor Imagery Classification Using Deep Residual Convolutional Networks***
Jing-Shan Huang, Wan-Shan Liu, Bin Yao, Zhan-Xiang Wang, Si-Fang Chen and Wei-Fang Sun

471 ***Vector Auto-Regressive Deep Neural Network: A Data-Driven Deep Learning-Based Directed Functional Connectivity Estimation Toolbox***
Takuto Okuno and Alexander Woodward for the Alzheimer's Disease Neuroimaging Initiative

492 ***Noise Robustness Low-Rank Learning Algorithm for Electroencephalogram Signal Classification***
Ming Gao, Runmin Liu and Jie Mao

501 ***Deep Convolutional Neural Network With a Multi-Scale Attention Feature Fusion Module for Segmentation of Multimodal Brain Tumor***
Xueqin He, Wenjie Xu, Jane Yang, Jianyao Mao, Sifang Chen and Zhanxiang Wang

510 ***Electroencephalogram Source Imaging and Brain Network Based Natural Grasps Decoding***
Baoguo Xu, Leying Deng, Dalin Zhang, Muhui Xue, Huijun Li, Hong Zeng and Aiguo Song

522 ***Quantitative Susceptibility Mapping-Derived Radiomic Features in Discriminating Multiple Sclerosis From Neuromyelitis Optica Spectrum Disorder***
Zichun Yan, Huan Liu, Xiaoya Chen, Qiao Zheng, Chun Zeng, Yineng Zheng, Shuang Ding, Yuling Peng and Yongmei Li

530 ***Frontal Cortex Segmentation of Brain PET Imaging Using Deep Neural Networks***
Qianyi Zhan, Yuanyuan Liu, Yuan Liu and Wei Hu

537 ***F-Value Time-Frequency Analysis: Between-Within Variance Analysis***
Hong Gi Yeom and Hyundoo Jeong

546 ***3D U-Net Improves Automatic Brain Extraction for Isotropic Rat Brain Magnetic Resonance Imaging Data***
Li-Ming Hsu, Shuai Wang, Lindsay Walton, Tzu-Wen Winnie Wang, Sung-Ho Lee and Yen-Yu Ian Shih

557 ***Discriminative Analysis of Schizophrenia Patients Using Topological Properties of Structural and Functional Brain Networks: A Multimodal Magnetic Resonance Imaging Study***
Jing Wang, Pengfei Ke, Jinyu Zang, Fengchun Wu and Kai Wu

570 ***Potential of Overcomplete Wavelet Frame Expansion for Facilitating Electroencephalogram Information Mining***
Wanshan Liu, Xiaoyue Guo, Binqiang Chen and Wangpeng He

575 ***Sparse Representation-Based Discriminative Metric Learning for Brain MRI Image Retrieval***
Guohua Zhou, Bing Lu, Xuelong Hu and Tongguang Ni

584 ***Continuous Monitoring of Mental Load During Virtual Simulator Training for Laparoscopic Surgery Reflects Laparoscopic Dexterity: A Comparative Study Using a Novel Wireless Device***
Neta B. Maimon, Maxim Bez, Denis Drobot, Lior Molcho, Nathan Intrator, Eli Kakiashvilli and Amitai Bickel

601 ***A Novel Constrained Non-negative Matrix Factorization Method for Group Functional Magnetic Resonance Imaging Data Analysis of Adult Attention-Deficit/Hyperactivity Disorder***
Ying Li, Weiming Zeng, Yuhu Shi, Jin Deng, Weifang Nie, Sizhe Luo and Jiajun Yang

614 ***Abnormal Cerebrovascular Reactivity and Functional Connectivity Caused by White Matter Hyperintensity Contribute to Cognitive Decline***
Dan Yang, Ruomeng Qin, Lan Chu, Hengheng Xu, Ling Ni, Junyi Ma, Pengfei Shao, Lili Huang, Bing Zhang, Meijuan Zhang and Yun Xu

625 ***AT-NeuroEAE: A Joint Extraction Model of Events With Attributes for Research Sharing-Oriented Neuroimaging Provenance Construction***
Shaofu Lin, Zhe Xu, Ying Sheng, Lihong Chen and Jianhui Chen

645 ***Residual Block Based Nested U-Type Architecture for Multi-Modal Brain Tumor Image Segmentation***
Sirui Chen, Shengjie Zhao and Quan Lan

655 ***Cross-Attention and Deep Supervision UNet for Lesion Segmentation of Chronic Stroke***
Manjin Sheng, Wenjie Xu, Jane Yang and Zhongjie Chen

664 ***A Network Pharmacology Based Research on the Mechanism of Donepezil in Treating Alzheimer's Disease***
Lihua Liu, Yingying Zhu, Peng Fu and Jundong Yang

671 ***High-Resolution Digital Panorama of Multiple Structures in Whole Brain of Alzheimer's Disease Mice***
Xianzhen Yin, Xiaochuan Zhang, Jingjing Zhang, Weicheng Yang, Xian Sun, Haiyan Zhang, Zhaobing Gao and Hualiang Jiang

688 ***Robust Latent Multi-Source Adaptation for Encephalogram-Based Emotion Recognition***
Jianwen Tao, Yufang Dan, Di Zhou and Songsong He

705 ***Multi-Model Adaptation Learning With Possibilistic Clustering Assumption for EEG-Based Emotion Recognition***
Yufang Dan, Jianwen Tao and Di Zhou

718 ***A Radiomics Nomogram for Classifying Hematoma Entities in Acute Spontaneous Intracerebral Hemorrhage on Non-contrast-Enhanced Computed Tomography***
Jia Wang, Xing Xiong, Jing Ye, Yang Yang, Jie He, Juan Liu and Yi-Li Yin

727 ***Motion Artifacts Removal and Evaluation Techniques for Functional Near-infrared Spectroscopy Signals: A Review***
Ruisen Huang, Keum-Shik Hong, Dalin Yang and Guanghao Huang

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 13 January 2021
doi: 10.3389/fnins.2020.615435

LatLRR-FCNs: Latent Low-Rank Representation With Fully Convolutional Networks for Medical Image Fusion

Zhengyuan Xu[1,2†], Wentao Xiang[1†], Songsheng Zhu[1], Rui Zeng[3], Cesar Marquez-Chin[4], Zhen Chen[1], Xianqing Chen[5], Bin Liu[1] and Jianqing Li[1*]*

[1] The Key Laboratory of Clinical and Medical Engineering, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China, [2] The Department of Medical Engineering, Wannan Medical College, Wuhu, China, [3] The Brain and Mind Centre, The University of Sydney, Sydney, NSW, Australia, [4] The KITE Research Institute, Toronto Rehabilitation Institute-University Health Network, Toronto, ON, Canada, [5] The Department of Electrical Engineering, College of Engineering, Zhejiang Normal University, Jinhua, China

OPEN ACCESS

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Siyuan Lu,
University of Leicester,
United Kingdom
David Haynor,
University of Washington,
United States

***Correspondence:**
Jianqing Li
jqli@njmu.edu.cn
Bin Liu
liubin@njmu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 09 October 2020
Accepted: 07 December 2020
Published: 13 January 2021

Citation:
Xu Z, Xiang W, Zhu S, Zeng R, Marquez-Chin C, Chen Z, Chen X, Liu B and Li J (2021) LatLRR-FCNs: Latent Low-Rank Representation With Fully Convolutional Networks for Medical Image Fusion. Front. Neurosci. 14:615435. doi: 10.3389/fnins.2020.615435

Medical image fusion, which aims to derive complementary information from multi-modality medical images, plays an important role in many clinical applications, such as medical diagnostics and treatment. We propose the LatLRR-FCNs, which is a hybrid medical image fusion framework consisting of the latent low-rank representation (LatLRR) and the fully convolutional networks (FCNs). Specifically, the LatLRR module is used to decompose the multi-modality medical images into low-rank and saliency components, which can provide fine-grained details and preserve energies, respectively. The FCN module aims to preserve both global and local information by generating the weighting maps for each modality image. The final weighting map is obtained using the weighted local energy and the weighted sum of the eight-neighborhood-based modified Laplacian method. The fused low-rank component is generated by combining the low-rank components of each modality image according to the guidance provided by the final weighting map within pyramid-based fusion. A simple sum strategy is used for the saliency components. The usefulness and efficiency of the proposed framework are thoroughly evaluated on four medical image fusion tasks, including computed tomography (CT) and magnetic resonance (MR), T1- and T2-weighted MR, positron emission tomography and MR, and single-photon emission CT and MR. The results demonstrate that by leveraging the LatLRR for image detail extraction and the FCNs for global and local information description, we can achieve performance superior to the state-of-the-art methods in terms of both objective assessment and visual quality in some cases. Furthermore, our method has a competitive performance in terms of computational costs compared to other baselines.

Keywords: multi-modality medical image, latent low-rank representation, fully convolutional networks, medical image fusion, Laplacian pyramid

1. INTRODUCTION

Medical image fusion is a key technology that has been used extensively in clinical diagnosis and treatment planning (James and Dasarathy, 2014). Modern medical imaging techniques mainly include computed tomography (CT), magnetic resonance (MR), single-photon emission computed tomography (SPECT), and positron emission tomography (PET) (Walrand et al., 2017). CT has a high spatial and density resolution for dense structures (e.g., bones, implants), while MR has a high resolution for soft tissue (Wang et al., 2016) (e.g., muscle, tendon, and fascia). PET is an advanced nuclear medical examination technique that allows visualization of biomolecular metabolism, receptors, and neurotransmitter activity *in vivo*. SPECT is often applied to quantify images of the physiological and pathological changes of organs or tissues *in vitro*. Evaluating different perspectives of these imaging techniques reveals that they do, to an extent, complement each other (Walrand et al., 2017). Thus, medical image fusion can be utilized to combine different medical images and generate a new fusing image, providing the clinical information from each original image (Du et al., 2016; Huang et al., 2020).

To date, many medical image fusion studies have been reported (Toet, 1989; Li et al., 1995, 2013; Petrovic and Xydeas, 2004; Lewis et al., 2007; Zhang and Guo, 2009; Bhatnagar et al., 2015; Wang Q. et al., 2015; Geng et al., 2017; Zhao and Lu, 2017; Li H. et al., 2018; Manchanda and Sharma, 2018). Among them, multiscale transform (MST)-based methods are commonly used. The key point of MST-based fusion techniques is to decompose the original images into a multiscale transform domain (Li et al., 1995). Some fusion rule strategies can be utilized to merge the transformed coefficients, and the merged coefficients are employed to reconstruct the composite image. Note that the current literature indicates that the non-subsampled shearlet transform (NSST) and non-subsampled contourlet transform (NSCT) achieve the optimum performance in terms of image representation among MST-based methods (Anitha et al., 2015; Li Y. et al., 2018; Yin et al., 2018; Zhu et al., 2019). Zhu et al. used NSCT to decompose medical image pairs into low-pass and high-pass sub-bands, where a phase congruency rule was applied to fuse the high-pass sub-bands and a local Laplacian energy-based fusion rule was utilized for the low-pass sub-bands (Zhu et al., 2019). Later, Yin et al. introduced a novel framework in which the high-frequency coefficients were fused by a parameter-adaptive pulse coupled neural network (PA-PCNN), and the weighted local energy and the weighted sum of eight-neighborhood-based modified Laplacian were utilized to fuse low-frequency bands in the NSST domain (Yin et al., 2018). However, due to the nature of the transformation, MST-based (including NCST-based and NSST-based) fusion methods may not express and extract certain significant structures of source images properly without being sensitive to misregistration.

To address the misregistration problem in MST-based methods, sparse representation (SR) has emerged as another popular and powerful theory in the medical image fusion field (Liu and Wang, 2014; Liu et al., 2016, 2019; Fei et al., 2017). A typical SR-based medical image fusion method includes three basic steps: (1) a given dictionary is used to find the sparsest representation of source images; (2) some fusion rules are used to integrate the sparse representation coefficients; and (3) the integrated sparse representation coefficients and given dictionary are utilized to construct the fused image. For example, Liu and Wang (2014) proposed a novel adaptive sparse representation model for medical image fusion, where a set of more compact sub-dictionaries was learned to replace the single redundant dictionary in the traditional SR approach and achieved better results. Although the SR-based and extended methods are robust in terms of noise and misregistration to some extent, they cannot capture global information and suffer from significant energy loss.

In the field of medical image fusion, a key issue is to calculate a weight map since it reflects pixel activity information from different modality images, determining the quality of the final fused image. The weight map is calculated by two steps: activity level measurement and weight assignment. However, these two steps suffer from the robustness problem because traditional methods cannot deal with noise and misregistration well, as indicated in Liu et al. (2017). To improve the robustness of activity level measurement and weight assignment, Liu et al. (2017) introduced a deep learning fusion method with a simple multi-layer convolutional neural network (CNN) using the decision map and the medical image under the pyramid-based image fusion framework to reconstruct the fused medical image. While such a method achieves some success in specific medical image fusion tasks, this work may fail in multi-modal image fusion because the simple use of the CNN cannot extract fine-grained details efficiently.

To address the aforementioned challenges, we propose a novel hybrid medical image fusion framework with two principal elements (e.g., LatLRR and FCNs), inspired by Liu and Wang (2014) and Liu et al. (2017). The main contributions of this paper are as follows:

- The latent low-rank representation (LatLRR) is applied to decompose the medical image into low-rank (for extraction of details) and saliency components (for the preservation of energies).
- In the context of the low-rank component, to avoid the fixed-length feature vector from the final full connection layer and the information loss in the traditional CNN, three different FCNs (due to the nature of an input image of arbitrary size) are applied to produce a correspondingly-sized feature map with an efficient deconvolution layer (Guo et al., 2018), where a prediction is generated for each pixel and the spatial information in the original input image is retained. A sum strategy is used to fuse the saliency parts for energy preservation.
- To the best of our knowledge, this fusion strategy in combination with LatLRR and FCNs is the first to be applied in the medical image domain.

The remainder of this paper is structured as follows. In section 2, the proposed fusion strategy is described in detail. Section

3 gives the experimental configurations. Section 4 illustrates a comparative study between the proposed frameworks and five representative medical image fusion methods in terms of visual quality and quantitative and computational cost assessments. The conclusion is drawn in section 5.

2. METHODOLOGY

As shown in **Figure 1**, each proposed framework is fed with a pair of pre-registered multi-modality medical source images, and outputs the fused medical image via the following four steps:

- We use the **LatLRR** theory to decompose the two medical source images into low-rank and saliency components (see section 2.1).
- To capture the detailed information of each source, a novel **fusion framework of the low-rank components** for each paired source based on FCNs, score maps, weight maps, and pyramid fusion is described (see section 2.2).
- To retain the energies of each source, a simple sum strategy is used to **fuse the saliency components** and **reconstruct the fused image** (see section 2.3).

2.1. LatLRR Decomposition

The LatLRR theory was first proposed by Liu and Yan (2011), integrating subspace segmentation and feature extraction simultaneously, to extract the global and local structure from raw data in the context of natural images. It can be summarized into the following problem (Li and Wu, 2018):

$$\begin{aligned} &\min_{X,Y,Z} \|X\|_* + \|Y\|_* + \lambda\|Z\|_1 \\ &s.t. Img = ImgX + ImgY + Z \end{aligned} \tag{1}$$

where $\|\|_*$ denotes the nuclear norm, $\|\|_1$ denotes the l_1-norm, and $\lambda > 0$ is the balance coefficient. *Img* is the observed data matrix, and X and Y denote the low-rank and saliency coefficients, respectively. Note that **Figure 2** explains the subject to Equation (1), where *ImgX*, *ImgY*, and Z are the low-rank, saliency, and noise components of *Img*, respectively.

In this paper, **the LatLRR decomposition** of Equation (1) can be solved by the inexact augmented Lagrangian multiplier method (Wang et al., 2013), where it extracts the low-rank and saliency components (e.g., $ImgX_j$ and $ImgY_j$) from medical image Img_j with $j = 1, 2$ (here, we consider two medical images, as shown in **Figure 1a**).

2.2. Fusion of Low-Rank Components

The **fusion of low-rank component** details can be seen in **Figure 1b**, including the FCN model for producing score maps (section 2.2.1), zero-phase component analysis (ZCA) (Kessy et al., 2018), and l_1-norm operations (section 2.2.2) for whiting the score maps and generating the weight maps, respectively, weighted local energy (WLE) and weighted sum of eight-neighborhood-based modified Laplacian (WSEML) (Yin et al., 2018) operations (section 2.2.3) for obtaining the fused weight map, and pyramid fusion strategy (section 2.2.4) for reconstructing the fused low-rank component.

2.2.1. FCN Model

The fully convolutional networks (FCNs), demonstrated in many studies (Long et al., 2015; Wang L. et al., 2015; Chen et al., 2017; Guo et al., 2018), achieved significant performance in image semantic segmentation. In the FCN architecture, after multiple convolutions and pooling processes, the obtained image size will be progressively smaller with a lower resolution, resulting in a heatmap (coarse output). To keep the output the same size as the input, a skip architecture is used for upsampling. In this work, three different scenarios are tested, as shown in **Figure 3**. For each scenario, there are 38 layers of FCNs before upsampling, including 16 convolutional layers (blue color block in **Figure 3A**), 15 rule layers, five pooling layers (green color block in **Figure 3A**), and two dropout layers. In **Figure 3A**, the FCN-32s is a single-stream net in which up-samples stride 32 predictions back to pixels in a single step, but the upsampling output is very coarse. To obtain the refined outputs of FCN-16s, the final layer and the pool4 layer are used to combine the predictions in **Figure 3B** at stride 16. In **Figure 3C**, to obtain the outputs of FCN-8s with greater precision, the pool3 layer, the pool4 layer, and the final layer are utilized to combine the predictions at stride 8. As shown in **Figure 1b**, the three trained FCNs (FCN-32s, FCN-16s, and FCN-8s) are utilized to classify a pair of the low-rank components of medical source images $Lr_1 = ImgX_1$ and $Lr_2 = ImgX_2$ pixel by pixel, producing the corresponding score maps ${S_j}^{1:C}, C = 21, j = 1, 2$ (the choice of $C = 21$ can be seen in section 3.4).

2.2.2. ZCA and l_1-Norm Operations

The details for ZCA and l_1-norm operations are depicted in **Figure 4**. To project the original redundancy score maps into a sparse subspace, we used ZCA to whiten those score maps ${S_j}^{1:C}$ and to obtain the score maps $\hat{S}_j^{1:C}$. Among the ZCA, the covariance matrix ${Co_j}^i$ is decomposed as follows:

$${Co_j}^i = {S_j}^i \times ({S_j}^i)^T, {Co_j}^i = U\Sigma V^T \tag{2}$$

where $i = 1, 2, \cdots, C; j = 1, 2$, and i denote the $i - th$ channel score map. Note that U, Σ and V define the left singular, singular values, and right singular matrixes, respectively (Chen et al., 2018). An alternative solution named $\hat{S}_j^i$ is given as follows:

$$\hat{S}_j^i = {K_j}^i \times {S_j}^i, {K_j}^i = U(\Sigma + \eta I)^{-\frac{1}{2}} U^T \tag{3}$$

where η is a small value avoiding bad matrix inversion and I is the identity matrix. Then, the local l_1-norm and average operations are used to calculate the initial weight map W_j:

$$W_j = \frac{\sum_{x=u-k}^{u+k} \sum_{y=v-k}^{v+k} \left\| \hat{S}_j^i(x, y) \right\|_1}{(2k+1) \times (2k+1)} \tag{4}$$

where $k = 2$ and the average l_1-norm is calculated by a window centered at $\hat{S}_j^i(u, v)$.

2.2.3. WLE and WSEML Operations

Once the initial weight maps W_1 and W_2 is calculated, the WLE and WSEML are applied to acquire the final fused weight map F_w, which is described in **Figure 1b** with the orange block.

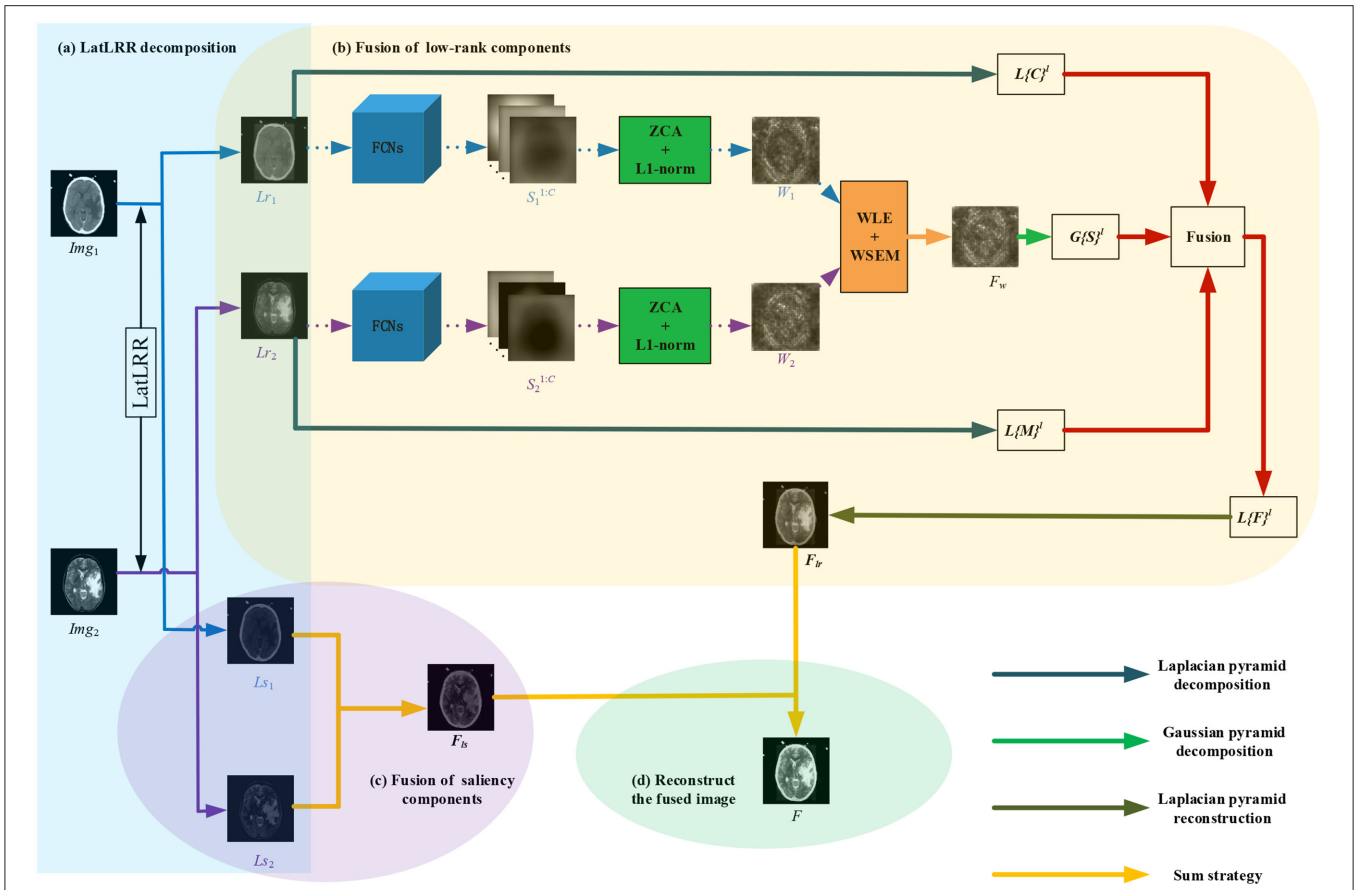


FIGURE 1 | Schematic diagram of the proposed end-to-end frameworks (LatLRR-FCNs). The proposed LatLRR-FCNs enable the fusion image to extract details and preserve energies from paired sources. It is composed of four parts: **(a)** LatLRR decomposition, **(b)** fusion of low-rank components, **(c)** fusion of saliency components, and **(d)** reconstruction of fused image. Img_1 and Img_2 are the source medical images, Lr_1 and Lr_2 are the low-rank components of Img_1 and Img_2, Ls_1 and Ls_2 are the saliency components of Img_1 and Img_2, $S_1^{1:C}$ and $S_2^{1:C}$ are the score maps, W_1 and W_1 are the initial weight maps of Lr_1 and Lr_2, and the final fused weight map is F_w. F_{lr} is the fused low-rank component, F_{ls} is the fused saliency component, and the final fused image is F.

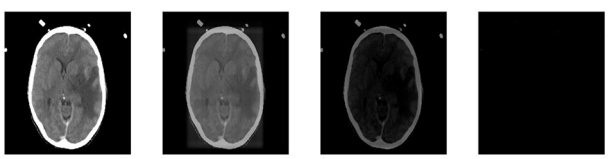


FIGURE 2 | The LatLRR decomposing operation. *Img* is the observed image. *ImgX* and *ImgY* are the low-rank and saliency components of *Img*, respectively. *Z* denotes the noisy component.

First, the WLE of each W_j (i.e., Φ_j) is calculated as follows:

$$\Phi_j(u,v) = \sum_{p=-r}^{r}\sum_{q=-r}^{r}\{\Omega(p+r+1, q+r+1)W_j(u+p, v+q)^2\} \tag{5}$$

where $j \in \{1,2\}$ and Ω denote a $(2r+1)\times(2r+1)$ weighting matrix. The value of each element in Ω is 2^{2r-d} with radius r, d denotes the element of a four-neighborhood distance to the center. If r is 1, Ω is equal to

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

.

Second, the WSEML of each W_j (i.e., Ψ_j) is given as follows:

$$\Psi_j(u,v) = \sum_{p=-r}^{r}\sum_{q=-r}^{r}\{\Omega(p+r+1, q+r+1) \times EML_j(u+p, v+q)\} \tag{6}$$

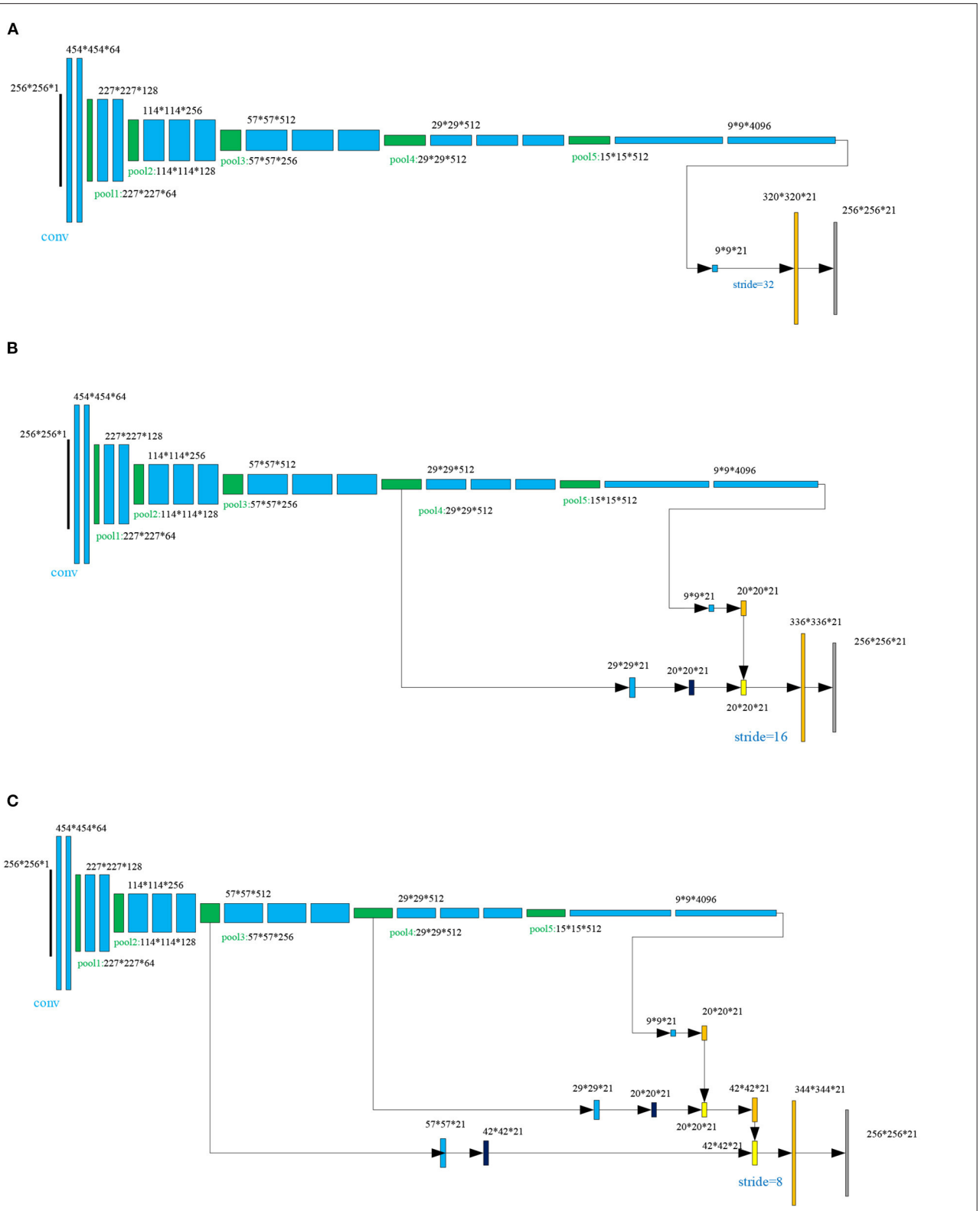


FIGURE 3 | The skip architecture for upsampling for three scenarios: **(A)** FNC-32s, **(B)** FNC-16s, and **(C)** FNC-8s.

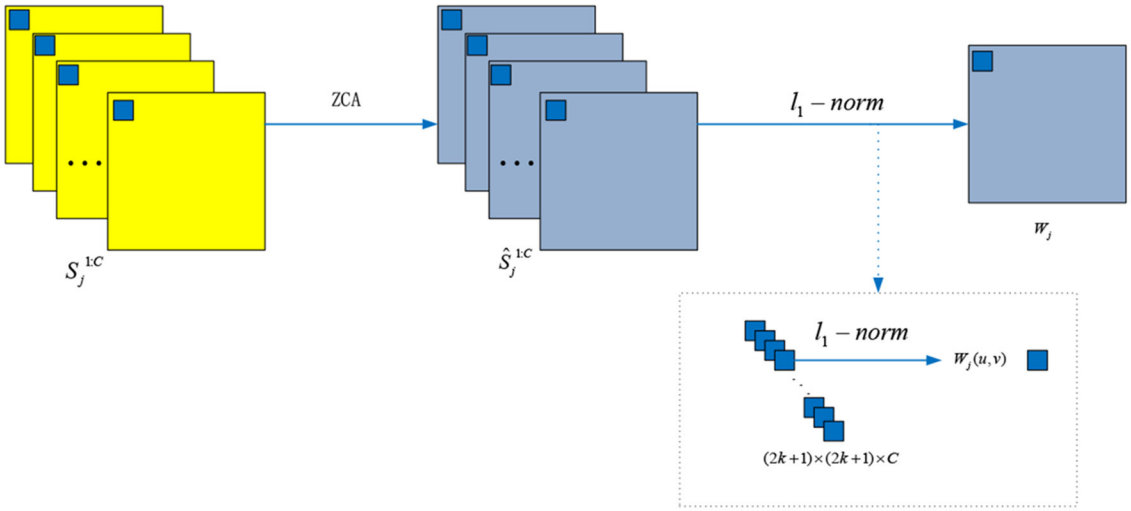


FIGURE 4 | The ZCA and l_1-norm operations for whiting the score maps and generating the weight maps, respectively.

where *EML* is expressed as follows:

$$\begin{aligned} EML_j(u,v) &= \left|2W_j(u,v) - W_j(u-1,v) - W_j(u+1,v)\right| \\ &+ \left|2W_j(u,v) - W_j(u,v-1) - W_j(u,v+1)\right| \\ &+\frac{1}{\sqrt{2}}\left|2W_j(u,v) - W_j(u-1,v-1) - W_j(u+1,v+1)\right| \\ &+\frac{1}{\sqrt{2}}\left|2W_j(u,v) - W_j(u-1,v+1) - W_j(u+1,v-1)\right| \end{aligned} \tag{7}$$

Finally, the fused weight map F_w is calculated by the following rule:

$$F_w(u,v) = \begin{cases} W_1(u,v), & \text{if } \Phi_1(u,v)\cdot\Psi_1(u,v) \geq \Phi_2(u,v)\cdot\Psi_2(u,v) \\ W_2(u,v), & \text{otherwise} \end{cases} \tag{8}$$

2.2.4. Pyramid Fusion Strategy

As shown in **Figure 1b**, the fused weight map F_w is decomposed into a Gaussian pyramid $G\{S\}^l$ (green color arrow). The low-rank components Lr_1 and Lr_2 are decomposed into a Laplacian pyramid (dark blue color arrow) $L\{C\}^l$ and $L\{M\}^l$, respectively. Note that l denotes the l-th decomposition level, which is calculated by the following:

$$l = \lfloor \log_2(X,Y) \rfloor \tag{9}$$

where $\lfloor\cdot\rfloor$ is the flooring operation and the spatial size of the low-rank component is $X \times Y$.

Next, those coefficients about $L\{F\}$ are calculated at each decomposition level l:

$$L\{F\}^l(u,v) = \begin{cases} \{G\{S\}^l(u,v)\cdot L\{C\}^l(u,v) \\ +(1-G\{S\}^l(u,v))L\{M\}^l(u,v)\}, & if\{Q^l(u,v) \geq \tau\} \\ L\{C\}^l(u,v), & if\{Q^l(u,v) < \tau \,\&\, E_C^l(u,v) \geq E_M^l(u,v)\} \\ L\{M\}^l(u,v), & if\{Q^l(u,v) < \tau \,\&\, E_C^l(u,v) \geq E_M^l(u,v)\} \end{cases} \tag{10}$$

where the threshold τ determines the corresponding fusion mode. $Q^l(x,y)$ is given as follows:

$$Q^l(u,v) = \frac{2\sum_p\sum_q L\{C\}^l(u+p,v+q)L\{M\}^l(u+p,v+q)}{E_C^l(u,v) + E_M^l(u,v)} \tag{11}$$

where $E_C^l(u,v)$ and $E_M^l(u,v)$ are the local energy maps of $L\{C\}^l$ and $L\{M\}^l$, respectively. $E_C^l(u,v)$ and $E_M^l(u,v)$ are defined as follows:

$$\begin{aligned} E_C^l(u,v) &= \sum_p\sum_q L\{C\}^l(u+p,v+q)^2 \\ E_M^l(u,v) &= \sum_p\sum_q L\{M\}^l(u+p,v+q)^2 \end{aligned} \tag{12}$$

Finally, the Laplacian pyramid reconstruction method (Mertens et al., 2009) (bottle green color arrow in **Figure 1b**) is used to reconstruct the fused low-rank components F_{lr} from $L\{F\}^l$, as indicated in Equation (10).

2.3. The Flowchart of the Proposed LatLRR-FCNs

The FCN architectures (FCN-32s or FCN-16s or FCN-8s) are inserted to produce two score maps with the focus property after the LatLRR decomposition once a pair of low-rank components for two images are calculated (hereafter, we named the proposed

Algorithm 1: LatRR-FCN-32s, LatRR-FCN-16s, and LatRR-FCN-8s networks.

Input: Medical images: Img_1 and Img_2.
Part 1: LatLRR decomposition.
  Parameters: the λ of LatLRR is set to 0.8.
for $j = 1:2$ **do**
  Perform the LatLRR decomposition on Img_j to obtain the low-rank component $Lr_j = ImgX_j$
  and the saliency component $Ls_j = ImgY_j$;
end
Part 2: Fusion of low-rank components.
  (i) FCN models
for $j = 1:2$ **do**
  The trained FCN (FCN-32s or FCN-16s or FCN-8s) models are used to
  process Lr_j to obtain $S_j^{1:C}, C = 21$;
end
  (ii) ZCA and l_1 -norm operations:
    Equations (2–4) are used to obtain W_1 and W_2;
  (iii) WLE and WSEML operations:
    Equations (5–8) are used to obtain F_w;
if $\Phi_1(u,v) \cdot \Psi_1(u,v) \geq \Phi_2(u,v) \cdot \Psi_2(u,v)$ **then**
  $F_w(u,v) = S_1(u,v)$;
else
  $F_w(u,v) = S_2(u,v)$;
end
  (iv) Pyramid fusion strategy;
    Decompose F_w into a Gaussian pyramid to obtain $G\{S\}^l$;
    Decompose Lr_1 and Lr_2 into a Laplacian pyramid to obtain $L\{C\}^l$ and $L\{M\}^l$, respectively;
    Equation (10) is used to obtain $L\{F\}^l$;
    Laplacian pyramid reconstruction is used to fuse F_{lr} from $L\{F\}^l$;
Part 3: Fusion of saliency components.
  $F_{ls} = Ls_1 + Ls_2$;
Part 4: Reconstruct the fused image.
  $F = F_{lr} + F_{ls}$
Output: the fused image F.

LatRR-FCNs: including proposed LatRR-FCN-32s, LatRR-FCN-16s and LatRR-FCN-8s, respectively). Algorithm 1 provides a pseudo-code of the proposed LatRR-FCN-32s, LatRR-FCN-16s and LatRR-FCN-8s networks. Then, ZCA and l_1-norm are utilized to white the score maps and obtain the initial weight maps for the low-rank components of paired source images [see **Part 2-(ii)** in Algorithm 1]. The WLE and WSEML techniques are used to fuse the two initial weight maps [see **Part 2-(iii)**]. The fused weight map and a pair of low-rank components under the pyramid-based image fusion framework (Mertens et al., 2009) are used to reconstruct the fused low-rank components' image F_{lr} [see **Part 2-(iv)**]. We sum the saliency components to obtain the fused saliency components' image F_{ls} (see **Part 3**). Finally, the fused image F is obtained by combining F_{lr} and F_{ls} (see **Part 4**).

3. EXPERIMENTAL CONFIGURATIONS

3.1. FCN Training Sets

Currently, transfer learning (Bar et al., 2015; Liu et al., 2017; Razzak et al., 2018; Lu et al., 2019, 2020) has become an active topic in the field of medical image analysis. In this study, we directly adopted a transfer learning strategy, and we trained the FCNs (FCN-32s, FCN-16s, and FCN-8s) on the PASCAL VOC 2012 dataset (Everingham et al., 2012) and the semantic boundary dataset (SBD) (Hariharan et al., 2011). The PASCAL VOC 2012 dataset contains 20 foreground object classes and 1 background class. The original dataset contains 1,464 (train), 1,449 (val), and 1,456 (test) pixel-level annotated images. The dataset is augmented with the SBD by extra annotations (Mertens et al., 2009), resulting in 10,582 training images.

3.2. Source Medical Image Testing Sets

In our experiments, we used 40 pairs of multi-modal medical images (each medical image fusion problem contains 10 image pairs) to demonstrate the usefulness and efficiency of the proposed methods. Most of the test images were gathered from the Whole Brain Atlas databases (Vidoni, 2012) and have been widely adopted in previous related publications (Liu and Wang, 2014; Liu et al., 2017, 2019; Yin et al., 2018; Zhu et al., 2019). Each pair of images was geometrically aligned, and all the test images were normalized to 256 × 256.

3.3. State-of-the-Art Methods

Five superior medical image fusion methods were collected for comparison against our proposed methods. These included the adaptive sparse representation (ASR) method (Liu and Wang, 2014) (https://github.com/yuliu316316/MST-SR-Fusion-Toolbox), the convolutional neural network (CNN)-based (LP-CNN) method (Liu et al., 2017) (https://github.com/yuliu316316/CNN-Fusion), the phase congruency and local Laplacian energy-based NSCT (NSCT-PC-LLE) method (Zhu et al., 2019) (https://github.com/zhiqinzhu123/Source-code-of-medical-image-fusion-in-NSCT-domain), the parameter-adaptive pulse coupled-neural network (NSST-PAPCNN) in the NSST domain method (Yin et al., 2018) (https://github.com/yuliu316316/NSST-PAPCNN-Fusion), and the convolutional sparsity-based morphological component analysis (CSMCA) method (Liu et al., 2019) (https://github.com/yuliu316316/CSMCA-Fusion). Among them, the NSCT-PC-LLE, NSST-PAPCNN, and CSMCA methods were proposed in last year.

3.4. Parameter Choices

The parameters of all compared methods were set to the default values. The key parameters for our proposed algorithms were given in **Table 1**. According to this table, the parameter λ in LatLRR decomposition was 0.8 (Li and Wu, 2018), and the threshold τ in Equation (10) was set to 0.8 (Liu et al., 2017). The PASCAL VOC 2012 dataset contained 20 foreground object classes and one background class, so that the C in $S_j^{1:C}$ was equal to 21. Note that we adopted a transfer learning strategy directly to train the FCN-VGG16 (Long et al., 2015) by MacInnes, and the trained models were obtained after 50 epochs using the

TABLE 1 | The key parameters used in our algorithms.

LatLRR decomposition	Threshold	Classes	No. of epochs
λ	τ	C	*epoch*
0.8	0.8	21	50

training data. The choice of epoch was dependent on **Figure 5**. When the epoch was lower than 50, the accuracy of the training and validation sets increased with the values of epochs. However, the accuracy of the validation set leveled off when the epoch was higher than 50, although the accuracy of the training set still increased regardless of the scenario (FCN-32s, FCN-16s, and FCN-8s). In terms of the loss function, the values for all FCN architectures decreased with the epoch in the case of the training set, but the loss of the scenarios tended to converge at the 50 epochs. Therefore, to balance the computational complexity and accuracy, the epochs for FCN models in this paper were chosen as 50.

3.5. Experimental Environment

All the experiments were implemented in MATLAB R2019a on a WIN64 Intel(R) Core (TM)i7-8750H CPU@2.20GHz 8GB RAM. The training models of the proposed method were trained in MATLAB R2019a+VS2017+ MatConvNet 1.0-beta25.

3.6. Objective Evaluation Metrices

In this study, five common representative quantitative metrics, e.g., *EN* (Liu and Wang, 2014), Q_{MI} (Bhatnagar et al., 2013), $Q^{AB/F}$ (Xydeas et al., 2000), *SCD* (Aslantas and Bendes, 2015), and *VIFF*(Han et al., 2013) (for all metrics, a larger value indicates a better performance), were used to evaluate the quality of fused images. The metrics were briefly described as follows:

(i) Entropy (*EN*) Liu and Wang (2014), Wang et al. (2017), Zhang et al. (2017): Entropy measures the amount of information in the fused image.

(ii) Mutual information (MI) of two images Q_{MI} Bhatnagar et al. (2013): MI is a quantitative assessment of the information shared by two images. Mathematically, MI can be expressed with joint entropy $H(C, D)$, marginal entropy $H(C)$, and $H(D)$ of two variables C and D as follows:

$$MI(C, D) = H(C) + H(D) - H(C, D) \tag{13}$$

where $H(C) = -\sum_{c} p(c)\log_2 p(c), H(D) = -\sum_{d} p(d)\log_2 p(d), H(C, D) = -\sum_{c,d} p(c, d)\log_2 p(c, d)$. $p(c)$ and $p(d)$ denote the marginal probability distributions of C and D, respectively. $p(c, d)$ denotes the joint probability distribution of C and D. Therefore, the quality of the fused image with respect to input images Img_1 and Img_2 can be defined as:

$$Q_{MI} = 2\left[\frac{MI(Img_1, F)}{H(Img_1) + H(F)} + \frac{MI(Img_2, F)}{H(Img_2) + H(F)}\right] \tag{14}$$

(iii) Edge-based similarity measure $Q^{AB/F}$: The authors in Xydeas et al. (2000) proposed a metric $Q^{AB/F}$ to produce the similarity between the edges that transform in the fusion process. This metric is defined as follows:

$$Q^{AB/F} = \frac{\sum_{u=1}^{N}\sum_{v=1}^{M}(Q^{AF}(u, v)w^{A}(u, v) + Q^{BF}(u, v)w^{B}(u, v))}{\sum_{u=1}^{N}\sum_{v=1}^{M}(w^{A}(u, v) + w^{B}(u, v))} \tag{15}$$

where A, B, and F represent the two input images (Img_1 and Img_2) and fused images. The size of each image is $N \times M$, $Q^{AF}(u, v)$ and $Q^{BF}(u, v)$ are defined as follows:

$$\begin{aligned} Q^{AF}(u, v) &= Q_g^{AF}(u, v)Q_\alpha^{AF}(u, v) \\ Q^{BF}(u, v) &= Q_g^{BF}(u, v)Q_\alpha^{BF}(u, v) \end{aligned} \tag{16}$$

where $Q_g^{*F}(u, v)$ and $Q_\alpha^{*F}(u, v)$ are the edge strength and orient preservation values at location (u, v) in images A and B, respectively. The dynamic range for $Q^{AB/F}$ is equal to $[0, 1]$, where a larger value for $Q^{AB/F}$ indicates a better fusion result. For more details of this metric, please refer to Xydeas et al. (2000).

(iv) The sum of the correlations of differences (SCD) Aslantas and Bendes (2015) is a quality metric formulated as follows:

$$SCD = r(D_1, Img_1) + r(D_2, Img_2) \tag{17}$$

where $D_1 = F - Img_2$, $D_2 = F - Img_1$, F is the fused image, and Img_1 and Img_2 are the input images. The $r(.)$ function calculates the correlation between S_k and D_k, given as:

$$r(D_k, Img_k) = \frac{\sum_{u}\sum_{v}(D_k(u, v) - \bar{D}_k)(Img_k(u, v) - \bar{Img}_k)}{\sqrt{(\sum_{u}\sum_{v} j(D_k(u, v) - \bar{D}_k)^2)\sum_{u}\sum_{v}(Img_k(u, v) - \bar{Img}_k)^2}} \tag{18}$$

where $k = 1, 2$, $\bar{D}_k$ and $\bar{Img}_k$ are the average of the pixel values of D_k and Img_k, respectively.

(v) The human visual perception-based metric visual information fidelity fusion (*VIFF*) Han et al. (2013): To obtain the *VIFF*, four steps are needed. First, the source and fused images are filtered and then divided into blocks. Second, visual information is evaluated with and without distortion information in each block. Third, the *VIFF* of each sub-band is calculated. Finally, the overall quality measure is determined by weighting the *VIFF* of each sub-band.

3.7. Color Space Fusion

In our proposed methods, the YUV color space was used to solve the grayscale and RGB color image (PET, SPECT) fusion issues. First, the RGB color image was converted into a YUV color space, resulting in three channel components of Y, U, and V. Then, the grayscale image and the Y channel were fused by using the proposed fusion methods, as described in section 2. Finally, the fused Y-channel component, the U-channel component, and the V-channel component were inversely transformed by YUV space, obtaining the fused color image.

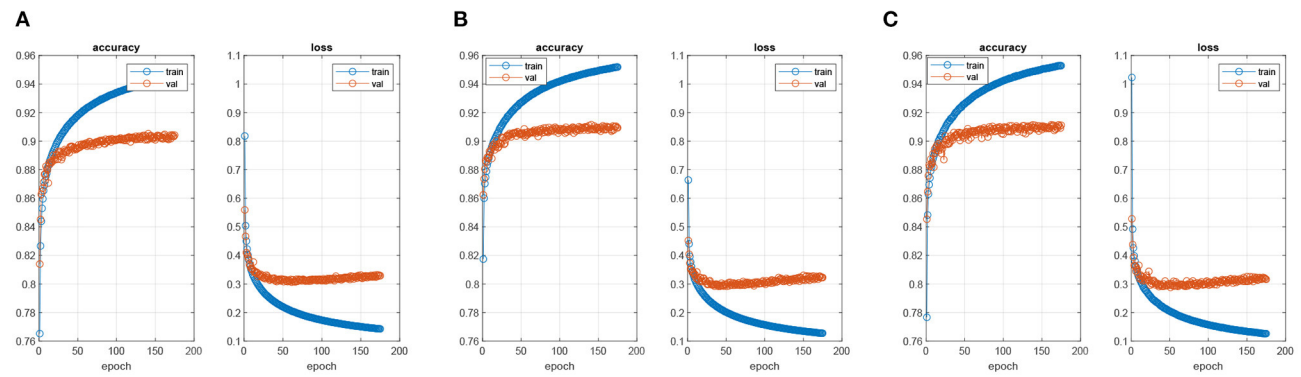


FIGURE 5 | The process of training FCN models. **(A)** FNC-32s model, **(B)** FNC-16s model, and **(C)** FNC-8s model.

4. RESULTS AND DISCUSSION

This section is devoted to showing that the proposed LatRR-FCNs can improve the information details and energy preservation in terms of **visual quality assessment** (section 4.1), **quantitative assessment** (section 4.2) and **computational cost assessment** (section 4.3), compared with five recently proposed methods: ASR (Liu and Wang, 2014), LP-CNN (Liu et al., 2017), NSCT-PC-LLE (Zhu et al., 2019), NSST-PAPCNN (Yin et al., 2018), and CSMCA (Liu et al., 2019). In this study, the usefulness and efficiency of each method are investigated with four sets of medical image fusion studies, including CT and MR, MR-T1 and MR-T2, PET and MR, and SPECT and MR.

4.1. Visual Quality

The fusion examples of CT and MR images are given in **Figure 6**. Furthermore, one representative region of each result is enlarged for better comparison. The ASR and CSMCA methods reveal a significant energy loss in both the CT and MR images (resulting in an intensity and contrast decrease in the fused images), especially for the bone and lesion regions in the **Figures 6a3–c3,a7–c7**. The fusion results of the NSCT-PC-LLE, LP-CNN, NSST-PAPCNN, and the proposed methods have better information preservation for the CT and MR modalities. However, the NSCT-PC-LLE, LP-CNN, and NSST-PAPCNN methods cannot extract the detailed information well in the MR image, which can be seen in the **Figures 6a4–c4,a5–c5,a6–c6** and the corresponding highlighted close-ups. Furthermore, the ASR method fails to extract the structural and edge details from the CT modality (see **Figures 6a4–c4**). The NSCT-PC-LLE and NSST-PAPCNN methods outperform the ASR method, even though some structural details cannot be extracted (see the **Figures 6a3–c3,a4–c4,a6–c6**). The proposed frameworks and LP-CNN method can effectively extract the structure and edge details from both CT and MR modalities (see **Figures 6a8–a10,b8–b10,c8–c10,a5–c5**, respectively). The proposed methods perform well on the preservation of detailed and structural information for all three examples.

Figure 7 gives three fusion examples of MR-T1 and MR-T2 images. The ASR and CSMCA methods suffer from low intensity and contrast caused by the loss of energy (see the **Figures 7a3–c3,a7–c7** with the close-up). In addition, the NSCT-PC-LLE, LP-CNN, and NSST-PAPCNN methods cannot preserve the detailed information (see the close-ups in **Figures 7a4–c4,a5–c5,a6–c6**, respectively). Furthermore, the ASR and NSCT-PC-LLE methods exhibit lower ability in structure and edge detail extraction within the MR-T1 modality, explained by the close-up in **Figures 7a4–c4,a5–c5**. Finally, compared to the other tested methods, our proposed LatLRR-FCN-based methods achieve the best performance, as shown with the close-ups in **Figures 7a8–c8,a9–c9,a10–c10**, respectively.

Figure 8 shows the three fusion examples of MR and PET images. The ASR and CSMCA methods lose a significant amount of energy in both the MR and PET modalities, as viewed in the **Figures 8a3–c3,a7–c7** and the corresponding close-ups. Note that the NSCT-PC-LLE and LP-CNN methods are subjected to a severe color distortion (see the close-ups in **Figures 8a4–c4,a5–c5**). Furthermore, the color distortion existed more or less in the fusion results of the NSST-PAPCNN method (see **Figures 8a7–c7** and the close-ups). Overall, the color preservation of our proposed algorithms (see **Figures 8a8–c8,a9–c9,a10–c10** together with their close-ups) are also significantly higher than the other methods.

The fusion examples of three sets of MR and SPECT images are shown in **Figure 9**. The ASR and CSMCA methods still lose much energy in both the MR and PET modalities (see **Figures 9a3–c3,a7–c7**). Moreover, color distortion exists in the NSCT-PC-LLE and LP-CNN methods (see the close-up in **Figures 9a4–c4,a5–c5**). Furthermore, in the results of the NSST-PAPCNN method, color distortion also exists (in **Figures 9a6–c6**, especially the close-up). The visual quality of color preservation of our proposed methods significantly outperforms the others.

4.2. Quantitative Assessment

Here, five common quantitative metrics as described in section 3.6 are employed to appraise the fusion performance. The average

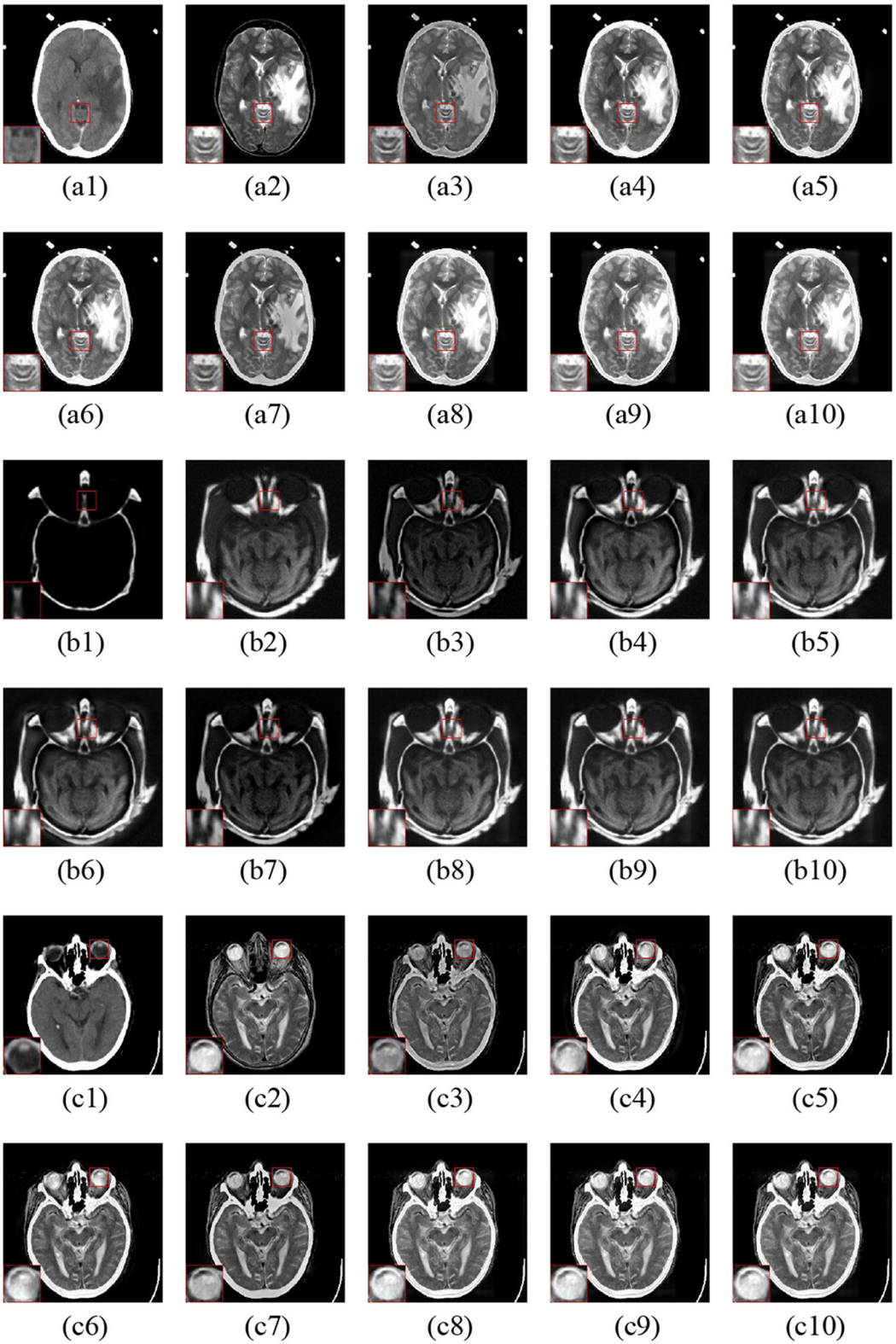


FIGURE 6 | Three fusion examples of CT and MR images. One close-up is provided in each set for better comparison. The original images: **(a1–c1)** CT and **(a2–c2)** MR. The fusion results **(a3–c3)** ASR, **(a4–c4)** NSCT-PC-LLE, **(a5–c5)** LP-CNN, **(a6–c6)** NSST-PAPCNN, **(a7–c7)** CSMCA, **(a8–c8)** LatLRR-FCN-32s, **(a9–c9)** LatLRR-FCN-16s, and **(a10–c10)** LatLRR-FCN-8s.

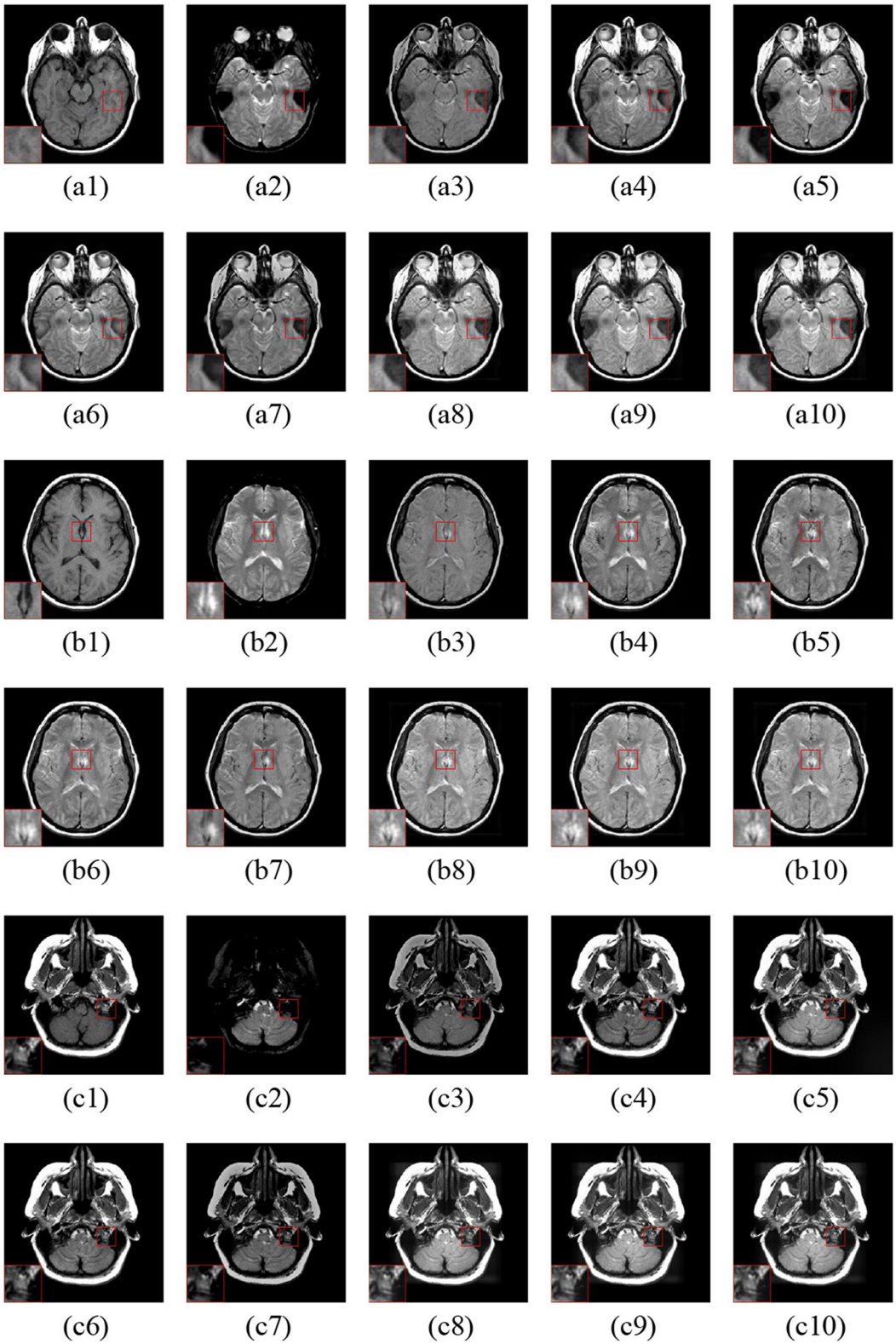


FIGURE 7 | Three fusion examples of MR-T1 and MR-T2 images. One close-up is provided in each set for better comparison. The original images: **(a1–c1)** MR-T1 and **(a2–c2)** MR-T2. The fusion results are as follows: **(a3–c3)** ASR, **(a4–c4)** NSCT-PC-LLE, **(a5–c5)** LP-CNN, **(a6–c6)** NSST-PAPCNN, **(a7–c7)** CSMCA, **(A8–C8)** LatLRR-FCN-32s, **(a9–c9)** LatLRR-FCN-16s, and **(a10–c10)** LatLRR-FCN-8s.

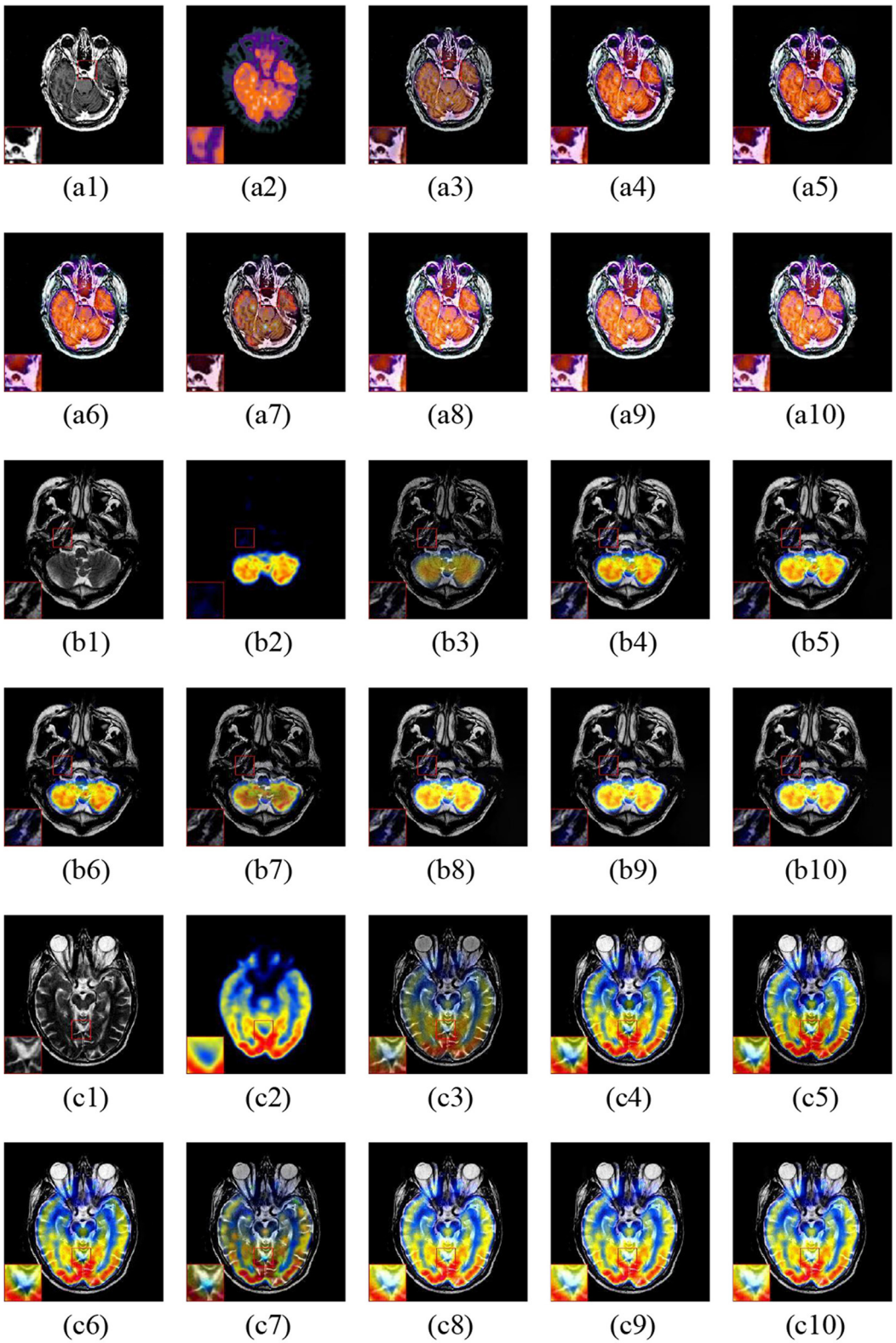


FIGURE 8 | Three fusion examples of MR and PET images. One close-up is provided in each set for better comparison. The original images: **(a1–c1)** MR and **(a2–c2)** PET. The fusion results: **(a3–c3)** ASR, **(a4–c4)** NSCT-PC-LLE, **(a5–c5)** LP-CNN, **(a6–c6)** NSST-PAPCNN, **(a7–c7)** CSMCA, **(a8–c8)** LatLRR-FCN-32s, **(a9–c9)** LatLRR-FCN-16s, and **(a10–c10)** LatLRR-FCN-8s.

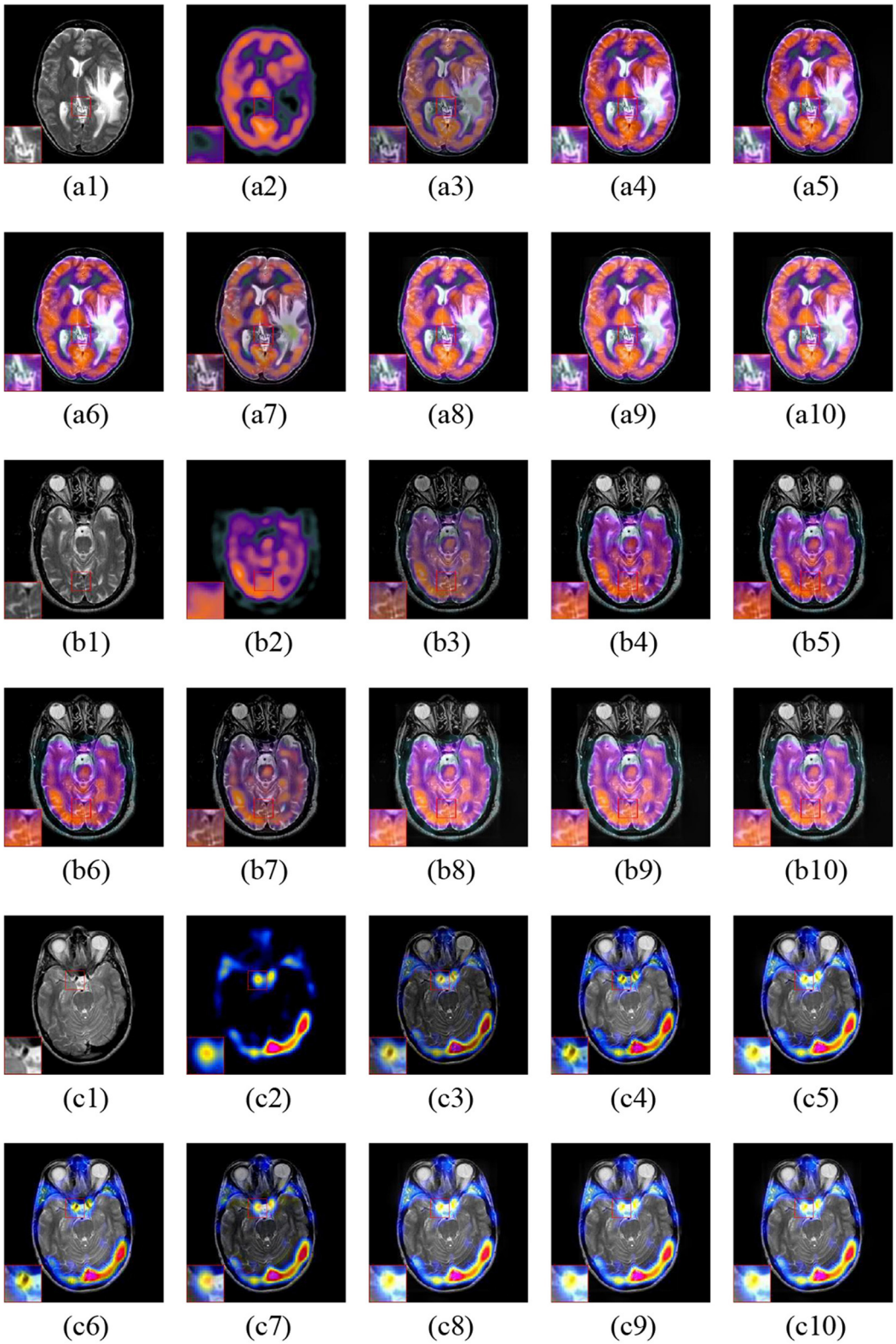


FIGURE 9 | Three fusion examples of MR and SPECT images. One close-up is provided in each set for better comparison. The original images: **(a1–c1)** MR and **(a2–c2)** SPECT. The fusion results: **(a3–c3)** ASR, **(a4–c4)** NSCT-PC-LLE, **(a5–c5)** LP-CNN, **(a6–c6)** NSST-PAPCNN, **(a7–c7)** CSMCA, **(a8–c8)** LatLRR-FCN-32s, **(a9–c9)** LatLRR-FCN-16s, and **(a10–c10)** LatLRR-FCN-8s.

TABLE 2 | Five metrics of different methods for the four medical image fusion problems (a higher value for each metric indicates a better performance).

CT and MR	***EN***	Q_{MI}	$Q^{AB/F}$	***SCD***	***VIFF***
ASR	4.7319	1.1068	0.5378	1.5400	0.3588
NSCT-PC-LLE	5.1619	1.1930	0.5206	1.5786	0.4570
LP-CNN	4.9503	1.1674	**0.5688**^{1}	1.5715	0.4476
NSST-PAPCNN	5.2323	1.1687	0.5383	1.6259	0.4703
CSMCA	4.8532	1.1120	**0.5535**^{3}	1.5462	0.4135
LatLRR-FCN-32s	**5.3915**^{3}	**1.2072**^{1}	**0.5541**^{2}	**1.6946**^{3}	**0.4998**^{3}
LatLRR-FCN-16s	**5.4037**^{2}	**1.2065**^{2}	0.5531	**1.6986**^{2}	**0.5003**^{2}
LatLRR-FCN-8s	**5.4531**^{1}	**1.2043**^{3}	0.5517	**1.7034**^{1}	**0.5007**^{1}
MR-T1 and MR-T2	***EN***	Q_{MI}	$Q^{AB/F}$	***SCD***	***VIFF***
ASR	4.0307	0.9965	0.5916	1.4834	0.4553
NSCT-PC-LLE	4.5478	**1.1500**^{1}	**0.6432**^{3}	1.4883	0.5934
LP-CNN	4.6327	**1.1294**^{3}	**0.6670**^{1}	1.5333	0.6006
NSST-PAPCNN	4.6654	**1.1305**^{2}	0.6035	1.5694	0.6053
CSMCA	4.1134	1.0401	**0.6606**^{2}	1.4677	0.5421
LatLRR-FCN-32s	**5.0259**^{3}	1.0940	0.6380	**1.7483**^{1}	**0.6335**^{3}
LatLRR-FCN-16s	**5.0524**^{2}	1.0953	0.6384	**1.7475**^{2}	**0.6346**^{2}
LatLRR-FCN-8s	**5.0949**^{1}	1.0974	0.6388	**1.7457**^{3}	**0.6358**^{1}
PET and MR	***EN***	Q_{MI}	$Q^{AB/F}$	***SCD***	***VIFF***
ASR	3.9649	0.9899	0.6153	1.6446	0.4206
NSCT-PC-LLE	4.4502	1.0739	0.6547	1.7013	0.5789
LP-CNN	4.5261	1.0523	0.6540	1.7121	0.5828
NSST-PAPCNN	4.4790	**1.0946**^{1}	0.6486	1.7398	0.5997
CSMCA	4.0233	1.0078	0.6573	1.6725	0.5015
LatLRR-FCN-32s	**4.5648**^{1}	1.0808	**0.6588**^{1}	**1.8861**^{1}	**0.6699**^{1}
LatLRR-FCN-16s	**4.5556**^{3}	**1.0816**^{3}	**0.6587**^{2}	**1.8851**^{2}	**0.6693**^{2}
LatLRR-FCN-8s	**4.5573**^{2}	**1.0819**^{2}	**0.6584**^{3}	**1.8834**^{3}	**0.6689**^{3}
SPECT and MR	***EN***	Q_{MI}	$Q^{AB/F}$	***SCD***	***VIFF***
ASR	4.5089	1.1563	0.5548	1.3995	0.4692
NSCT-PC-LLE	4.9146	**1.3703**^{2}	**0.6405**^{2}	1.3839	0.5629
LP-CNN	5.4404	**1.3146**^{3}	**0.6496**^{1}	1.4958	0.6024
NSST-PAPCNN	4.9242	**1.3912**^{1}	**0.6355**^{3}	1.5119	0.5814
CSMCA	4.6187	1.1977	0.6233	1.4073	0.5339
LatLRR-FCN-32s	**5.6361**^{1}	1.2880	0.6205	**1.8224**^{3}	**0.7075**^{3}
LatLRR-FCN-16s	**5.6270**^{3}	1.2899	0.6207	**1.8244**^{2}	**0.7078**^{2}
LatLRR-FCN-8s	**5.6341**^{2}	1.2916	0.6210	**1.8265**^{1}	**0.7081**^{1}

score of each method in each fusion problem is reported in **Table 2**. The top three values of all the fusion methods are shown in bold, and their rank is indicated by a superscript. For CT and MRI fusion, the proposed methods achieve the best performance in terms of *EN* (i.e., the values of *EN* for LatLRR-FCN-8s, LatLRR-FCN-16s, and LatLRR-FCN-32s are equal to **5.4531**^{1}, **5.4037**^{2} and **5.3915**^{3}, respectively), Q_{MI} (the value of Q_{MI} for LatLRR-FCN-32s, **1.2072**^{1}, higher than that of LatLRR-FCN-16s, **1.2065**^{2}, and LatLRR-FCN-8s, **1.2043**^{3}), *SCD*, and *VIFF* metrics. Note that in the context of the $Q^{AB/F}$ metric, our proposed LatLRR-FCN-32s ($Q^{AB/F} = \mathbf{0.5541}^{2}$) is slightly lower than the LP-CNN method ($Q^{AB/F} = \mathbf{0.5688}^{1}$) but slightly superior to the CSMCA method ($Q^{AB/F} = \mathbf{0.5535}^{3}$). In the case of MR-T1 and MR-T2 fusion, the proposed methods show the best values in three of the five metrics with *EN*, *SCD* and *VIFF*. Among them, an increase improvement in the proposed LatLRR-FCN-32s for *SCD* about **10.23**% [(1.7483 − 1.5694)/1.7483 = 0.1023] is reported in **Table 2**, compared to the best performance among the other five methods, i.e., NSST-PAPCNN algorithm. For MRI and PET fusion, overall, the proposed LatRR-FCNs obtain the best results in all five objective metrics except that the NSST-PAPCNN method achieves the rank first in the metric $Q^{AB/F}$ (i.e., $Q^{AB/F} = \mathbf{1.0946}^{1}$) with a slight improvement **1.16**% (e.g., 0.0116 = (1.0946 − 1.0819)/1.0946) compared with our proposed LatLRR-FCN-8s. Finally, our proposed LatRR-FCNs outperform the other fusion methods in the aspect of *EN*, *SCD*, and *VIFF* metrics for MRI and SPECT fusion, especially for the *SCD* metric of LatRR-FCN-8s with a significant improvement in **17.22**% (0.1722 = (1.8265 − 1.5119)/1.8265) compared to that of the NSST-PAPCNN approach.

As also shown in **Table 2**, for different metrics, it can be concluded as follows. (1) For the *EN* metric, our proposed techniques have the optimal energy preservation in four medical image fusion problems. (2) The Q_{MI} metric shows that our proposed LatLRR-FCN-8s and LatLRR-FCN-16s architectures obtain the best performance in detail information extraction than others in the context of CT and MR image fusion and PET and MR image fusion problems. (3) In terms of the $Q^{AB/F}$ metric, our proposed frameworks are also close to the other comparison algorithms in edge and direction retention. (4) For the *SCD* metric, our proposed methods have a higher cross-correlation between the fused image and the input image than the others in all four medical image fusion problems. (5) For the *VIFF* metric, compared to the other methods, our proposed approaches are more consistent with the visual mechanism of human eyes in four medical image fusion problems.

Moreover, **Figure 10** shows the objective performance of different methods in each fusion problem. The ten scores of each method in each fusion problem are connected for each metric. Obviously, the proposed three methods show the optimal performance among them. More specifically, the proposed LatLRR-FCNs are the best three ranks on the metrics of *EN*, *SCD*, and *VIFF* for all four problems, which is also concluded in **Table 2**.

4.3. Computational Cost Assessment

The average computational costs of different methods are shown in **Table 3**, including gray-level and color images. Although the performances of LP-CNN, NSCT-PC-LLE, and NSST-PAPCNN are better than the proposed methods, the proposed methods achieve a better performance in terms of both visual perception and objective assessment. However, the processing cost of ASR and CSMCA is 6 times and 10 times higher than our proposed methods. In total, the experimental results show that the proposed methods can

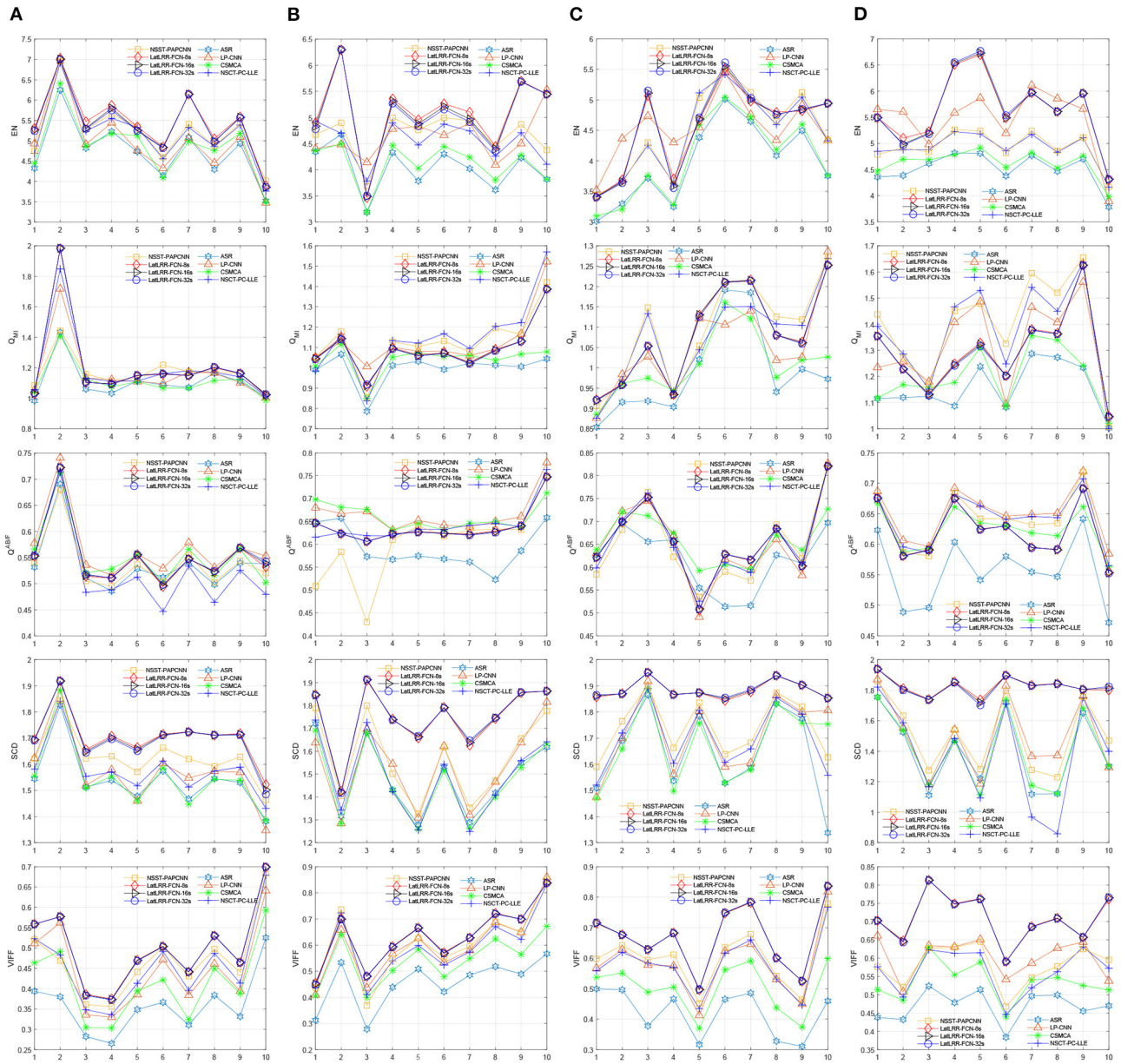


FIGURE 10 | Objective performance of different fusion methods in each fusion problem. **(A)** CT and MR, **(B)** MR-T1 and MR-T2, **(C)** PET and MR, and **(D)** SPECT and MR.

achieve competitive performance in terms of computational costs in practice.

5. CONCLUSION

In this paper, three LatRR-FCNs have been proposed to improve energy conservation and detail extraction during medical image fusion. Based on LatLRR, the LatRR-FCNs decompose the medical image into low-rank and saliency components, which can enhance the extraction of detail in the SR-based methods. Then, three different fully convolutional networks (FCN-32s, FCN-16s, and FCN-8s), ZCA, l_1-norm, WLE, and WSEML operations together with a pyramid-based fusion method are applied to fuse the low-rank components, which can simultaneously enhance the energy preservation and detail extraction. We sum the saliency components to obtain the fused saliency components. Finally, the fused image is obtained by combining the fused low-rank components and fused saliency components. The proposed frameworks were evaluated in the context of four kinds of medical image

TABLE 3 | Computation cost of different methods.

Methods	Gray-level image fusion time (s)	Color image fusion time (s)
ASR	128.2032	163.4961
NSCT-PC-LLE	2.1638	4.2115
LP-CNN	11.7130	11.9583
NSST-PAPCNN	4.2226	4.8991
CSMCA	233.4387	252.1574
LatLRR-FCN-32s	21.5983	19.5076
LatLRR-FCN-16s	21.7452	19.7250
LatLRR-FCN-8s	20.2011	18.7253

fusion problems, including CT and MR, MR-T1 and MR-T2, PET and MR, and SPECT and MR. The results of our experiments demonstrated that the proposed frameworks can achieve optimal performance in both visual quality and objective assessment.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.med.harvard.edu/aanlib/.

AUTHOR CONTRIBUTIONS

ZX and JL: conceptualization. ZX and WX: methodology and writing–review and editing. ZX: software, visualization, and writing–original draft preparation. SZ, CM-C, and ZC: validation. RZ and XC: formal analysis. JL: resources and supervision. BL: data curation and project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This document is the results of the research project funded by the National Key Research and Development Programme of China (2017YFB1303200), NSFC (61571113, 62001240), Leading-edge Technology and Basic Research Programme of Jiangsu (BK20192004D), Jiangsu Graduate Scientific Research Innovation Programme (KYCX20_1391), and Zhejiang Provincial Natural Science Foundation of China under Grant (LY17F010003).

ACKNOWLEDGMENTS

The authors would like to thank the editor and reviewers for the valuable comments and suggestions. We would like to thank Yu Liu for providing the experimental codes for their methods.

REFERENCES

Anitha, S., Subhashini, T., and Kamaraju, M. (2015). A novel multimodal medical image fusion approach based on phase congruency and directive contrast in nsct domain. *Int. J. Comput. Appl.* 129, 30–35. doi: 10.5120/ijca2015907014

Aslantas, V., and Bendes, E. (2015). A new image quality metric for image fusion: the sum of the correlations of differences. *AEU Int. J. Electron. Commun.* 69, 1890–1896. doi: 10.1016/j.aeue.2015.09.004

Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., and Greenspan, H. (2015). "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)* (Brooklyn, NY: IEEE), 294–297. doi: 10.1109/ISBI.2015.7163871

Bhatnagar, G., Wu, Q. J., and Liu, Z. (2013). Directive contrast based multimodal medical image fusion in nsct domain. *IEEE Trans. Multimedia* 15, 1014–1024. doi: 10.1109/TMM.2013.2244870

Bhatnagar, G., Wu, Q. J., and Liu, Z. (2015). A new contrast based multimodal medical image fusion framework. *Neurocomputing* 157, 143–152. doi: 10.1016/j.neucom.2015.01.025

Chen, Q., Xu, J., and Koltun, V. (2017). "Fast image processing with fully-convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 2497–2506. doi: 10.1109/ICCV.2017.273

Chen, Y., Yang, M., Chen, X., Liu, B., Wang, H., and Wang, S. (2018). Sensorineural hearing loss detection via discrete wavelet transform and principal component analysis combined with generalized eigenvalue proximal support vector machine and tikhonov regularization. *Multimedia Tools Appl.* 77, 3775–3793. doi: 10.1007/s11042-016-4087-6

Du, J., Li, W., Lu, K., and Xiao, B. (2016). An overview of multi-modal medical image fusion. *Neurocomputing* 215, 3–20. doi: 10.1016/j.neucom.2015.07.160

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2012). *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results*. Available online at: http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html

Fei, Y., Wei, G., and Zongxi, S. (2017). Medical image fusion based on feature extraction and sparse representation. *Int. J. Biomed. Imaging* 2017, 1–13. doi: 10.1155/2017/3020461

Geng, P., Sun, X., and Liu, J. (2017). Adopting quaternion wavelet transform to fuse multi-modal medical images. *J. Med. Biol. Eng.* 37, 230–239. doi: 10.1007/s40846-016-0200-6

Guo, X., Nie, R., Cao, J., Zhou, D., and Qian, W. (2018). Fully convolutional network-based multifocus image fusion. *Neural Comput.* 30, 1775–1800. doi: 10.1162/neco_a_01098

Han, Y., Cai, Y., Cao, Y., and Xu, X. (2013). A new image fusion performance metric based on visual information fidelity. *Inform. Fusion* 14, 127–135. doi: 10.1016/j.inffus.2011.08.002

Hariharan, B., Arbeláez, P., Bourdev, L., Maji, S., and Malik, J. (2011). "Semantic contours from inverse detectors," in *2011 International Conference on Computer Vision* (Barcelona: IEEE), 991–998. doi: 10.1109/ICCV.2011.6126343

Huang, B., Yang, F., Yin, M., Mo, X., and Zhong, C. (2020). A review of multimodal medical image fusion techniques. *Comput. Math. Methods Med.* 2020, 1–16. doi: 10.1155/2020/8279342

James, A. P., and Dasarathy, B. V. (2014). Medical image fusion: a survey of the state of the art. *Inform. Fusion* 19:4–19. doi: 10.1016/j.inffus.2013.12.002

Kessy, A., Lewin, A., and Strimmer, K. (2018). Optimal whitening and decorrelation. *Am. Stat.* 72, 309–314. doi: 10.1080/00031305.2016.1277159

Lewis, J. J., O'Callaghan, R. J., Nikolov, S. G., Bull, D. R., and Canagarajah, N. (2007). Pixel-and region-based image fusion with complex wavelets. *Inform. Fusion* 8, 119–130. doi: 10.1016/j.inffus.2005.09.006

Li, H., He, X., Tao, D., Tang, Y., and Wang, R. (2018). Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recogn.* 79, 130–146. doi: 10.1016/j.patcog.2018.02.005

Li, H., Manjunath, B., and Mitra, S. K. (1995). Multisensor image fusion using the wavelet transform. *Graph. Models Image Process.* 57, 235–245. doi: 10.1006/gmip.1995.1022

Li, H., and Wu, X.-J. (2018). Infrared and visible image fusion using latent low-rank representation. *arXiv [Preprint]. arXiv:1804.08992*. Available online at: https://arxiv.org/abs/1804.08992

Li, S., Kang, X., and Hu, J. (2013). Image fusion with guided filtering. *IEEE Trans. Image Process*. 22, 2864–2875. doi: 10.1109/TIP.2013.2244222

Li, Y., Sun, Y., Huang, X., Qi, G., Zheng, M., and Zhu, Z. (2018). An image fusion method based on sparse representation and sum modified-laplacian in nsct domain. *Entropy* 20:522. doi: 10.3390/e20070522

Liu, G., and Yan, S. (2011). "Latent low-rank representation for subspace segmentation and feature extraction," in *2011 International Conference on Computer Vision* (Barcelona: IEEE), 1615–1622. doi: 10.1109/ICCV.2011.6126422

Liu, Y., Chen, X., Cheng, J., and Peng, H. (2017). "A medical image fusion method based on convolutional neural networks," in *2017 20th International Conference on Information Fusion (Fusion)* (Xi'an: IEEE), 1–7. doi: 10.23919/ICIF.2017.8009769

Liu, Y., Chen, X., Ward, R. K., and Wang, Z. J. (2016). Image fusion with convolutional sparse representation. *IEEE Signal Process. Lett*. 23, 1882–1886. doi: 10.1109/LSP.2016.2618776

Liu, Y., Chen, X., Ward, R. K., and Wang, Z. J. (2019). Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Process. Lett*. 26, 485–489. doi: 10.1109/LSP.2019.2895749

Liu, Y., and Wang, Z. (2014). Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Process*. 9, 347–357. doi: 10.1049/iet-ipr.2014.0311

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 3431–3440. doi: 10.1109/CVPR.2015.7298965

Lu, S., Lu, Z., and Zhang, Y.-D. (2019). Pathological brain detection based on alexnet and transfer learning. *J. Comput. Sci*. 30, 41–47. doi: 10.1016/j.jocs.2018.11.008

Lu, S., Wang, S.-H., and Zhang, Y.-D. (2020). Detection of abnormal brain in MRI via improved Alexnet and ELM optimized by chaotic bat algorithm. *Neural Comput. Appl*. 32, 1–13. doi: 10.1007/s00521-020-05082-4

Manchanda, M., and Sharma, R. (2018). An improved multimodal medical image fusion algorithm based on fuzzy transform. *J. Visual Commun. Image Represent*. 51, 76–94. doi: 10.1016/j.jvcir.2017.12.011

Mertens, T., Kautz, J., and Van Reeth, F. (2009). "Exposure fusion: a simple and practical alternative to high dynamic range photography," in *Computer Graphics Forum*, Vol. 28, eds R. Scopigno and E. Gröller (Wiley Online Library), 161–171. doi: 10.1111/j.1467-8659.2008.01171.x

Petrovic, V. S., and Xydeas, C. S. (2004). Gradient-based multiresolution image fusion. *IEEE Trans. Image Process*. 13, 228–237. doi: 10.1109/TIP.2004.823821

Razzak, M. I., Naz, S., and Zaib, A. (2018). "Deep learning for medical image processing: overview, challenges and the future," in *Classification in BioApps*, eds N. Dey, A. Ashour, and S. Borra (Cham: Springer), 323–350. doi: 10.1007/978-3-319-65981-7_12

Toet, A. (1989). A morphological pyramidal image decomposition. *Pattern Recogn. Lett*. 9, 255–261. doi: 10.1016/0167-8655(89)90004-4

Vidoni, E. D. (2012). The whole brain atlas: www.med.harvard.edu/aanlib/. *J. Neurol. Phys. Therapy* 36:108. doi: 10.1097/NPT.0b013e3182563795

Walrand, S., Hesse, M., and Jamar, F. (2017). "SPECT/CT, PET/CT and PET/MR principles," in *Diagnostic and Therapeutic Nuclear Medicine for Neuroendocrine Tumors*, eds K. Pacak and D. Taïeb (Cham: Springer), 163–200. doi: 10.1007/978-3-319-46038-3_8

Wang, L., Ouyang, W., Wang, X., and Lu, H. (2015). "Visual tracking with fully convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago), 3119–3127. doi: 10.1109/ICCV.2015.357

Wang, P., Zhang, C., Cai, S., and Li, L. (2013). Accelerated matrix recovery via random projection based on inexact augmented lagrange multiplier method. *Trans. Tianjin Univ*. 19, 293–299. doi: 10.1007/s12209-013-2135-0

Wang, Q., Li, S., Qin, H., and Hao, A. (2015). Robust multi-modal medical image fusion via anisotropic heat diffusion guided low-rank structural analysis. *Inform. Fusion* 26, 103–121. doi: 10.1016/j.inffus.2015.01.001

Wang, S., Yang, M., Du, S., Yang, J., Liu, B., Gorriz, J. M., et al. (2016). Wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in mri scanning. *Front. Comput. Neurosci*. 10:106. doi: 10.3389/fncom.2016.00106

Wang, S., Yang, M., Li, J., Wu, X., Wang, H., Liu, B., et al. (2017). Texture analysis method based on fractional fourier entropy and fitness-scaling adaptive genetic algorithm for detecting left-sided and right-sided sensorineural hearing loss. *Fundament. Inform*. 151, 505–521. doi: 10.3233/FI-2017-1507

Xydeas, C., and Petrovic, V. (2000). Objective image fusion performance measure. *Electron. Lett*. 36, 308–309. doi: 10.1049/el:20000267

Yin, M., Liu, X., Liu, Y., and Chen, X. (2018). Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Trans. Instrument. Meas*. 68, 49–64. doi: 10.1109/TIM.2018.2838778

Zhang, Q., and Guo, B. L. (2009). Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Process*. 89, 1334–1346. doi: 10.1016/j.sigpro.2009.01.012

Zhang, Y., Ranjan Nayak, D., Yang, M., Yuan, T.-F., Liu, B., Lu, H., et al. (2017). Detection of unilateral hearing loss by stationary wavelet entropy. *CNS Neurol. Disord. Drug Targets* 16, 122–128. doi: 10.2174/1871527315666161026115046

Zhao, W., and Lu, H. (2017). Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation. *IEEE Trans. Instrument. Meas*. 66, 2283–2294. doi: 10.1109/TIM.2017.2700198

Zhu, Z., Zheng, M., Qi, G., Wang, D., and Xiang, Y. (2019). A phase congruency and local laplacian energy based multi-modality medical image fusion method in nsct domain. *IEEE Access* 7, 20811–20824. doi: 10.1109/ACCESS.2019.2898111

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

The handling editor declared a past co-authorship with one of the authors XC.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 21 January 2021
doi: 10.3389/fnins.2020.620235

Edge-Aware Pyramidal Deformable Network for Unsupervised Registration of Brain MR Images

*Yiqin Cao[1], Zhenyu Zhu[1], Yi Rao[1], Chenchen Qin[2], Di Lin[3], Qi Dou[4], Dong Ni[1] and Yi Wang[1]**

[1] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, [2] Tencent AI Lab, Shenzhen, China, [3] The College of Intelligence and Computing, Tianjin University, Tianjin, China, [4] Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China

Deformable image registration is of essential important for clinical diagnosis, treatment planning, and surgical navigation. However, most existing registration solutions require separate rigid alignment before deformable registration, and may not well handle the large deformation circumstances. We propose a novel edge-aware pyramidal deformable network (referred as EPReg) for unsupervised volumetric registration. Specifically, we propose to fully exploit the useful complementary information from the multi-level feature pyramids to predict multi-scale displacement fields. Such coarse-to-fine estimation facilitates the progressive refinement of the predicted registration field, which enables our network to handle large deformations between volumetric data. In addition, we integrate edge information with the original images as dual-inputs, which enhances the texture structures of image content, to impel the proposed network pay extra attention to the edge-aware information for structure alignment. The efficacy of our EPReg was extensively evaluated on three public brain MRI datasets including Mindboggle101, LPBA40, and IXI30. Experiments demonstrate our EPReg consistently outperformed several cutting-edge methods with respect to the metrics of Dice index (DSC), Hausdorff distance (HD), and average symmetric surface distance (ASSD). The proposed EPReg is a general solution for the problem of deformable volumetric registration.

Keywords: deformable image registration, convolutional neural networks, brain MR image, affine registration, 3D registration

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Bing Li, Hong Kong Polytechnic University, Hong Kong
Marco Fusella, Veneto Institute of Oncology (IRCCS), Italy

***Correspondence:**
Yi Wang
onewang@szu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *22 October 2020*
Accepted: *28 December 2020*
Published: *21 January 2021*

Citation:
Cao Y, Zhu Z, Rao Y, Qin C, Lin D, Dou Q, Ni D and Wang Y (2021) Edge-Aware Pyramidal Deformable Network for Unsupervised Registration of Brain MR Images. Front. Neurosci. 14:620235. doi: 10.3389/fnins.2020.620235

1. INTRODUCTION

Deformable image registration is to perform spatial transformation between a pair of images and establish a non-linear point-wise correspondence to achieve spatial consistency (Sotiras et al., 2013). By doing so, mono-/multi-modality information can be fused into the same coordinate system. It plays a very important role in various medical imaging studies to provide complementary diagnostic information and investigate changes of anatomical structures. Although many algorithms have been proposed over the past few decades (Sotiras et al., 2013; Shen et al., 2017; Haskins et al., 2020), registration is still a challenging task. Traditional registration methods may be computationally expensive and time-consuming due to their iterative optimization during deformation estimation procedure. Moreover, most existing deformable registration solutions

require separate rigid alignment before non-rigid registration, and may not well handle the large deformation circumstances. Therefore, efficient and accurate deformable registration scheme is still greatly expected to compensate for complicated non-rigid deformations.

As illustrated in **Figure 1**, the goal of registration is to match all corresponding anatomical structures in two images to the same spatial system through plausible deformable transformation. To calculate the desired deformable transformation, several non-linear deformation algorithms (Klein et al., 2009) have been proposed, such as large diffeomorphic distance metric mapping (LDDMM) (Beg et al., 2005; Auzias et al., 2011), standard symmetric normalization (Avants et al., 2008), and Demons (Vercauteren et al., 2009). These methods treat deformable registration as a procedure of iterative optimization for maximizing the similarity [such as mean square error (MSE) (Wolberg and Zokai, 2000), normalized mutual information (NMI) (Knops et al., 2006), and normalized cross-correlation (NCC) (Rao et al., 2014), etc.] between fixed and warped moving images. However, the iterative optimization strategy may take a relatively long time to deal with complicated volumetric deformations.

To address aforementioned issue, deep neural networks have been widely investigated for the registration task in recent years (Haskins et al., 2020). The registration networks are beneficial to aggregate abundant features from paired images to effectively predict the deformation field. Eppenhof et al. (2018) employed synthetic random transformations to train a registration framework based on a convolutional neural network (CNN). Fan et al. (2019) also applied a supervised CNN for image registration by using obtained ground-truth deformation fields as the supervision information. Uzunova et al. (2017) synthesized a large amount of realistic ground-truth data using model-based strategy to train a registration network. Yang et al. (2017) proposed a patch-wise deformation prediction model, which is a deep encoder-decoder network devised to estimate the momentum-parameterization of LDDMM model. The major limitation of supervised registration networks (Cao et al., 2017; Sokooti et al., 2017; Uzunova et al., 2017; Yang et al., 2017; Eppenhof et al., 2018; Fan et al., 2019) is the prerequisite of the ground-truth registration fields, which would highly affect the network performance. However, unlike segmentation or detection tasks, it is always difficult to obtain registration ground-truth.

In contrast, some studies have focused on unsupervised deep learning algorithms which achieved great success in various registration tasks (Sheikhjafari et al., 2018; Kuang and Schmah, 2019). The mechanism of unsupervised registration networks is to build model to obtain the deformation fields based on maximizing the similarity between two images, thus is without the need of ground-truth deformations. Li and Fan (2018) proposed to predict deformation parameters using a fully convolutional network, but this is a 2D approach that tends to ignore the volumetric information. Rohé et al. (2017) utilized U-net (Ronneberger et al., 2015) to estimate the deformation field of 3D cardiac MR images and employed the sum of squared differences (SSD) as the similarity loss. Balakrishnan et al. (2018) proposed an end-to-end network with cross-correlation as its loss function and spatial transformer networks (STN) (Jaderberg et al., 2015) as warping module. However, the prerequisite of this network was another rigid alignment. In addition to unsupervised learning, weakly supervised registration methods usually pay extra attention on the correspondences between structural information of two images, such as the extracted corresponding anatomical landmarks in prostate MR and ultrasound images (Hu et al., 2018). The weakly supervised network with structural similarity could provide more reliable registration but still requires a small amount of manual annotations.

One major challenge facing existing registration neural networks is the effective solution for large deformation compensation. To tackle this issue, Hu et al. (2019) proposed a registration network based on (Balakrishnan et al., 2018), which warps the multi-resolution feature maps to obtain the deformation field. However, in such a way, the low-resolution deformation field cannot be accurately acted on subsequent high-resolution features, thus may degrade the registration accuracy. At the same time, many other cascade/recursive networks (de Vos et al., 2019; Zhao et al., 2019a,b) have been proposed. The general idea of these networks is to progressively estimate the complicated transformation relationship between moving and fixed images, which is similar to the iterative optimization idea of traditional algorithms. For example, deep learning image registration (DLIR) (de Vos et al., 2019) combined affine and non-linear networks to calculate both affine alignment and non-linear registration. Volume tweening network (VTN) (Zhao et al., 2019b) cascaded several registration sub-networks, which deforms the moving images by multiple times according to the deformation estimation. The recursive cascade network (Zhao et al., 2019a) expanded the number of cascaded networks, and only calculated the similarity of the last cascade for training. In general, the cascade/recursive networks simplify the challenge of large deformation based on progressive deformation estimation. But the performance of these networks would be affected by the training strategies and the cumulative errors caused by the cascaded propagation.

In this study, we devise a novel edge-aware pyramidal deformable network (EPReg) for unsupervised volumetric registration. The proposed EPReg is a dual-stream pyramid framework, which utilizes original images and corresponding edge-aware maps to compose dual inputs, and generates multi-scale paired feature maps for recursively transforming the information between images into more representative features to predict more accurate deformation field. Finally, the trained EPReg can perform deformable registration in one forward pass. Extensive experiments on three 3D brain magnetic resonance imaging (MRI) datasets demonstrate that our proposed network achieves satisfactory registration performance.

The main contributions of our work are 2-fold.

1. We propose to fully exploit the useful complementary information from the multi-level feature pyramids to predict multi-scale deformation fields. Such coarse-to-fine estimation facilitates the progressive refinement of the predicted

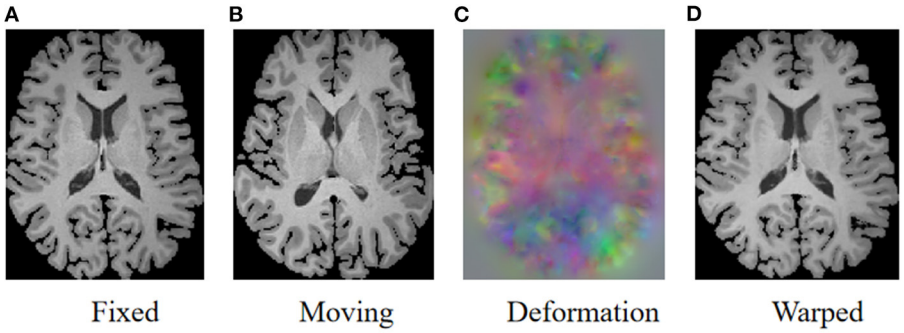


FIGURE 1 | Illustration of image registration. Given a fixed image **(A)** and a moving image **(B)**, a deformation field **(C)** is calculated to warp the moving image so that the warped image **(D)** is registered to the fixed image.

registration field, which enables our network to handle large deformations between volumetric images.
2. We integrate edge information with the original images as dual-inputs, which enhances the texture structures of image content, to impel the proposed network pay extra attention to the edge-aware information for structure alignment.

The remainder of this paper is organized as follows. Section 2 presents the details of the edge-aware pyramidal deformable network. Section 3 shows the experimental results of the proposed EPReg for the application of brain MRI registration. Section 4 elaborates the discussion of the proposed network, and the conclusion of this study is given in section 5.

2. EDGE-AWARE PYRAMIDAL DEFORMABLE NETWORK

The proposed registration network is illustrated in **Figure 2** (A) with its affine alignment block (B) and deformable registration block (C). We denote the input volumetric image pair as a fixed volume (I_f) and a moving volume (I_m). Their edge maps are denoted as E_f and E_m, respectively. The EPReg network leverages deformable pyramid to progressively transform the information between I_f-I_m and E_f-E_m into more representative features to predict more accurate deformation field ($\phi_4 \sim \phi_1$). The trained EPReg can attain deformable registration in one forward pass.

The following subsections first give a brief introduction on the deformable registration and then present the details of our scheme and elaborate the novel edge-aware pyramidal deformable architecture.

2.1. Preliminaries

Volumetric registration is to establish the voxel-wise correspondences between different volumes (i.e., fixed volume $I_f \in \mathbb{R}^3$ and moving volume $I_m \in \mathbb{R}^3$). The goal is to predict the optimal deformation field ϕ, so that the warped moving volume $I_m \circ \phi \in \mathbb{R}^3$ can be matched with I_f. The optimization problem can be defined as:

$$\phi = \arg\min_{\phi} \mathcal{L}_{sim}(I_f, I_m \circ \phi) + \lambda \mathcal{L}_{smooth}(\phi), \quad (1)$$

where $I_m \circ \phi$ denotes I_m warped by ϕ. $\mathcal{L}_{sim}$ defines similarity criterion and $\mathcal{L}_{smooth}$ regularizes the deformation ϕ to match any specific properties in the solution, and λ is a regularization parameter. There are several conventional formulations for $\mathcal{L}_{sim}$ and $\mathcal{L}_{smooth}$, respectively. Common similarity measures include MSE, NMI, NCC, and structural similarity index (SSIM) (Wang et al., 2004). $\mathcal{L}_{smooth}$ is often formulated as a regularizer on the spatial gradients of the displacement field.

2.2. Pyramidal Deformable Network

The proposed EPReg is build on the dual-stream pyramid architecture as shown in **Figure 2A**. The dual-stream encoder part is with shared parameters. As shown in **Figure 3**, the encoder part consists of four down-sampling convolutional blocks. Each convolutional block contains a 3D strided convolution with stride of 2, to reduce the spatial dimension in half. For the second and third convolutional blocks, residual connections (He et al., 2016) are employed. Specifically, two residual blocks are successively employed, each of which consists of two convolutional layers with a residual connection. For the last convolutional block, a 3D atrous spatial pyramid pooling (ASPP) (Wang et al., 2019) module is used to resample features at different scales for the capture of more representative multi-scale information. Batch normalization and rectified linear unit (ReLU) operations are applied in each convolutional block.

The convolutional blocks capture hierarchical paired features (i.e., $F_1 \sim F_4$ and corresponding $M_1 \sim M_4$) of the input volumetric pair, which are then used to progressively predict multi-scale deformation field ($\phi_4 \sim \phi_1$). Such coarse-to-fine estimation based on paired feature pyramids enhances the capability for handing large-scale deformation estimation.

We first perform rigid alignment on the feature maps F_4 and M_4 with high-level semantic information. Specifically, we devise an affine block to achieve global alignment. As shown in **Figure 2B**, the affine block consists of an affine convolutional layer (a residual block and a $1 \times 1 \times 1$ convolution operation) and a global average pooling (GAP) layer. It takes paired feature maps F_4-M_4 as input, and outputs the affine deformation field ϕ_4, which contains 12 degrees of freedom:

$$\phi_4 = f_a(F_4, M_4; \theta_a), \quad (2)$$

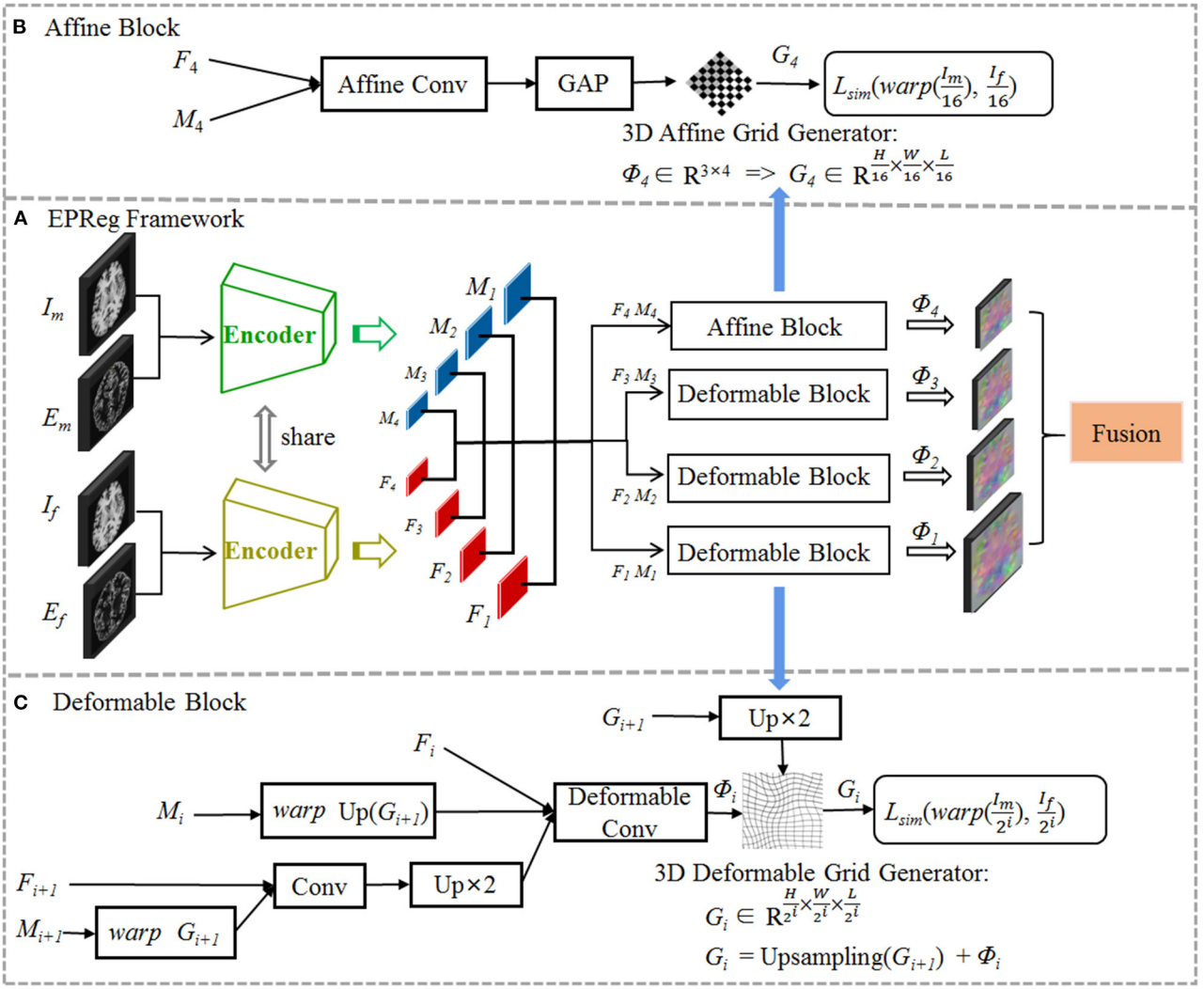


FIGURE 2 | The proposed edge-aware pyramidal deformable network for volumetric image registration. The schematic illustration of the proposed **(A)** dual-stream multi-level registration architecture, **(B)** affine alignment block, and **(C)** deformable registration block. The EPReg utilizes original images (fixed image I_f and moving image I_m) and their corresponding edge maps (E_f and E_m) to compose dual inputs, and generates multi-scale paired feature maps ($F_1 \sim F_4$ and $M_1 \sim M_4$) for transforming the information between images into more representative features to predict more accurate deformation field ($\phi_4 \sim \phi_1$).

where θ_a represents the parameter learned by affine block f_a. According to the estimated ϕ_4, the 3D affine grid G_4 can be generated and then warp the moving volume to rigidly align with the fixed volume.

Based on paired feature pyramids, we then progressively carry out the non-rigid registration via the devised deformable block (see **Figure 2C**). The deformable block contains a deformable convolutional layer, which consists of two residual blocks and a $1 \times 1 \times 1$ convolution operation. To estimate ϕ_i, the input of the deformable block includes three components, i.e., F_i, warped M_i using previously estimated ϕ_{i+1}, and the fused previous feature maps (F_{i+1} and warped M_{i+1} using ϕ_{i+1}):

$$\phi_i = f_{d_i}(F_i, M_i, F_{i+1}, M_{i+1}, \phi_{i+1}; \theta_{d_i}), \tag{3}$$

where θ_{d_i} represents the parameter learned by the i-th deformable block f_{d_i}, and $i = 1, 2, 3$. In such a way, based on previous deformation estimation (i.e., ϕ_{i+1}) and paired feature pyramids, each deformable block further estimates extra deformation ϕ_i, which can integrates with ϕ_{i+1} to attain more accurate non-rigid registration. For the i-th deformable block, the 3D deformable grid G_i is the combination of G_{i+1} (with 2×upsampling) and ϕ_i (see **Figure 2C**). **Figure 4** illustrates one example of the progressive deformation estimation and registration. It can be observed that the progressive deformation estimation can gradually refine the whole deformation field. The low-resolution deformation field ϕ_3 contains coarse and global deformation information, while the high-resolution deformation field ϕ_1 captures more detailed local displacements. Thus, the whole deformation field can attain accurate registration.

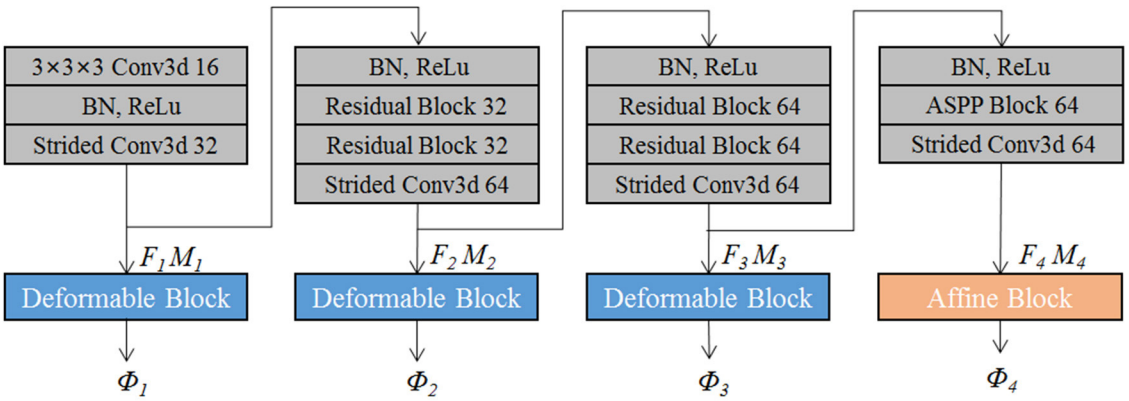


FIGURE 3 | The schematic illustration of the encoder architecture.

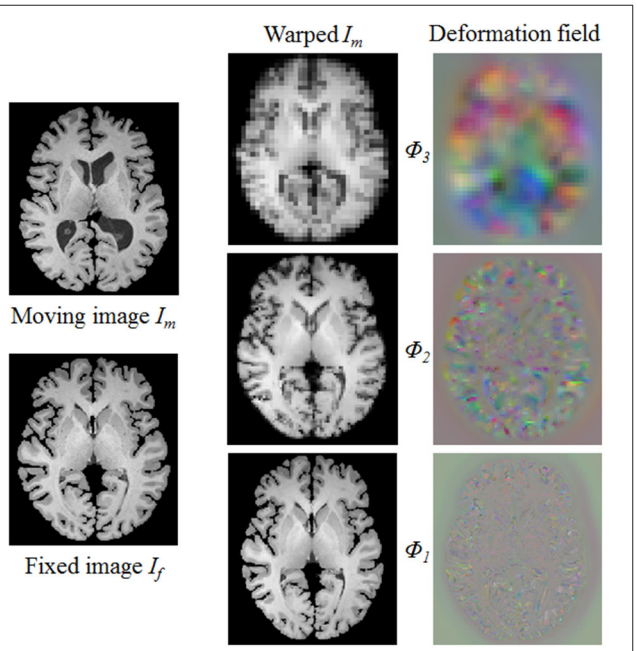


FIGURE 4 | One example to illustrate the multi-scale deformable registration fields and the procedure of progressive registration. It can be observed that the progressive deformation estimation can gradually generate more explicit local displacements, thus provide more and more accurate registration even for the large deformation case.

In summary, we utilize multi-level feature pyramids generated from paired volumes to estimate the multi-scale deformation fields. Our network generates the multi-scale deformation fields in a coarse-to-fine manner, which aggregates both high-level context information and low-level details. High-level context information is applied to the coarse and rigid alignment, while low-level details are devoted to the non-rigid registration.

2.3. Edge-Aware Dual Inputs

We integrate edge information with the original image as dual-inputs for each encoding stream (see **Figure 5**), which enhances the texture structures of image content, to impel the proposed network pay extra attention to the edge-aware information for structure alignment.

Considering the effectiveness and easy implementation of the Sobel edge detector, 3D Sobel operator is designed to extract the edges of original volume. The 3D Sobel operator contains three filtering kernels as S_x, S_y, and S_z. Each kernel is a $3 \times 3 \times 3$ tensor, and is responsible for the calculation of image gradient along x-/y-/z-axis. The kernel S_z is shown as an example:

$$S_z(:,:,-1) = \begin{bmatrix} +1 & +2 & +1 \\ +2 & +4 & +2 \\ +1 & +2 & +1 \end{bmatrix}, S_z(:,:,0) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ S_z(:,:,+1) = \begin{bmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{bmatrix}. \tag{4}$$

Kernels S_x and S_y are with the same kernel weights as S_z, but along different directions. S_x, S_y, and S_z are applied to convolve a volumetric image I, and further generate its corresponding edge map E as follows:

$$E = \sqrt{(S_x * I)^2 + (S_y * I)^2 + (S_z * I)^2}, \tag{5}$$

where $*$ denotes the convolution operation. The generated edge map is beneficial to impel the network leverage edge-aware information for structure alignment.

2.4. Training Loss

We adopt the patch-based cross-correlation (Rao et al., 2014) as the similarity function:

$$\mathcal{L}_{sim}(I_f, I_m) = \\ -\sum_{p \in \Omega} \frac{\left(\sum_{p_n}(I_f(p_n) - \bar{I}_f(p))(I_m(p_n) - \bar{I_m}(p))\right)^2}{\sum_{p_n}\left(I_f(p_n) - \bar{I}_f(p)\right)\sum_{p_n}\left(I_m(p_n) - \bar{I_m}(p)\right)}, \tag{6}$$

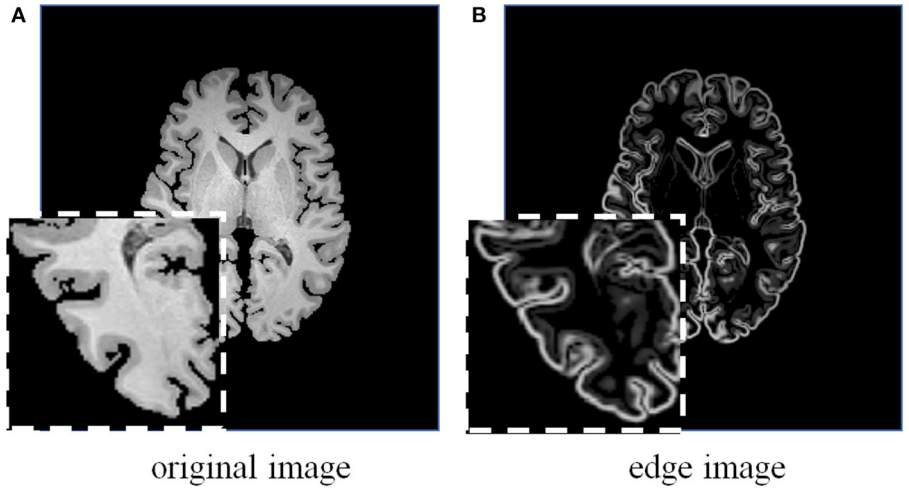


FIGURE 5 | A brain MRI image **(A)** and its corresponding edge map **(B)** generated using 3D Sobel operator. The edge map can more explicitly represent the structural information.

where $\bar{I_f}$ and $\bar{I_m}$ denote volumes with local mean intensities subtracted. $p \in \Omega$ denotes each voxel in volumetric image, where Ω is the whole image domain. Voxel p_n is the local neighborhood in v^3 ($v = 9$ in our implementation) volumetric patch at the center of voxel p.

To avoid obtaining an unpractical or discontinuous deformation field, we also add a diffusion regularizer $\mathcal{L}_{smooth}$ to impose smooth constraint on the spatial gradients of the overall deformation field ϕ:

$$\mathcal{L}_{smooth}(\phi) = \sum_{p \in \Omega} \left\| \nabla \phi \left(p \right) \right\|^2, \tag{7}$$

where $\phi = \sum_{i=1}^{4} upsample2^i(\phi_i)$ is the aggregation of multi-scale deformation fields.

As the deformation is estimated progressively, we consider similarity loss for each scale of the registration pyramid. Therefore, the total loss is defined as:

$$\mathcal{L} = \sum_{i=1}^{4} \mathcal{L}_{sim}(down2^i(I_f), down2^i(I_m) \circ G_i) + \lambda \mathcal{L}_{smooth}(\phi), \tag{8}$$

where $down2^i$ denotes a down-sample operation with a factor of 2^i. G_i is the 3D deformation grid generated to warp the moving volume.

3. EXPERIMENTS AND RESULTS

3.1. Materials

The study protocol was reviewed and approved by the Institution's Ethical Review Board. Experiments were carried on three public brain MRI datasets with manually labeled region of interests (ROIs), including Mindboggle101 (Klein and Tourville, 2012), LPBA40 (Shattuck et al., 2008), and IXI30 (Serag et al., 2012).

- Mindboggle101 (101 T1-weighted MRI volumes): 62 volumes were involved to conduct experiments as described in Kuang and Schmah (2019). Specifically, 42 volumes (i.e., 1,722 pairs) from subsets of NKI-RS-22 and NKI-TRT-20 were used for training, and 20 volumes (i.e., 380 pairs) from OASIS-TRT-20 were involved for testing.
- LPBA40 (40 T1-weighted MRI volumes): 30 volumes were randomly selected for training and the remaining 10 volumes were used for testing.
- IXI30 (30 T1-weighted MRI volumes): all 30 volumes were used for testing. In order to investigate the generalization ability of the network, we employed the model trained on LPBA40 to register images from IXI30 dataset.

All volumes were pre-processed by histogram and intensity normalization, and skull-stripping using FreeSurfer (Fischl, 2012).

3.2. Implementation Details

In our experiments, each input volumetric image was resized into the dimension of 192 × 192 × 192. The network was trained on a GPU of NVIDIA Tesla V100. The value of the regularization parameter λ was set empirically as 1000. For the whole registration network, the number of epochs was set to 300. The network was implemented using Pytorch and Adam optimization (Kingma and Ba, 2014), and the learning rate was initially set to 2e-4, with 0.5 weight decay after every 10 epoch.

3.3. Evaluation Metrics

To quantitatively evaluate the registration performance, Dice index (DSC) (Dice, 1945), Hausdorff distance (HD) (Huttenlocher et al., 1993), and average symmetric surface distance (ASSD) (Taha and Hanbury, 2015) were

calculated. The DSC is defined as:

$$DSC = \frac{2|R_{I_f} \cap R_{I_m}|}{|R_{I_f}| + |R_{I_m}|}, \tag{9}$$

where R_{I_f} and R_{I_m} are the segmented ROIs of I_f and I_m, respectively. The HD measures the longest distance over the shortest distances between the segmented ROIs of I_f and I_m. The ASSD is defined as:

$$ASSD = \frac{1}{|B_{I_f}| + |B_{I_m}|}\left(\sum_{x \in B_{I_f}} d(x, B_{I_m}) + \sum_{y \in B_{I_m}} d(y, B_{I_f})\right) \tag{10}$$

where B_{I_f} and B_{I_m} are the segmented surfaces of I_f and I_m, respectively. The operator $d(,)$ is the shortest Euclidean distance operator.

All evaluation metrics were calculated in 3D. A better registration shall have larger DSC, and smaller HD and ASSD.

3.4. Registration Accuracy

We compared our EPReg network with four cutting-edge brain MRI registration schemes: SyN (Avants et al., 2008), VoxelMorph (Balakrishnan et al., 2018), FAIM (Kuang and Schmah, 2019), and MSNet (Duan et al., 2019). For a fair comparison, we obtained their results either by directly taking the results from their papers or by generating the results from the public codes provided by the authors using the recommended parameter setting. In addition, we also compared the network without edge-aware input, which is denoted as PReg.

Figure 6 shows the visual comparisons from different registration methods on Mindboggle101 dataset. Our network can generate more accurate registered images, and the internal structures can be preserved consistently by using our network. **Table 1** further reports the numerical results on five regions of the images from dataset Mindboggle101. It can be observed that our EPReg consistently achieved best registration performance with respect to DSC and HD metrics. Regarding the ASSD evaluation, our network obtained the best ASSD values on occipital and temporal regions; and the second best ASSD on the frontal, parietal and cingulate regions. It is worth noting that both our EPReg and PReg networks outperformed other cutting-edge methods by a large margin in terms of DSC values, which demonstrates the proposed deformable pyramid contributed to the improvement of registrations.

For the LPBA40 dataset, the visualization and quantitative results (including seven regions) are shown in **Figure 7** and **Table 2**, respectively. It can be observed that our network achieved overall satisfactory registration performance on LPBA40 dataset. We further calculated DSC, HD, ASSD values of 54 corresponding sub-regions from warped and fixed volumes. The comparison results are illustrated in **Figure 8**. For the 54 sub-regions, the proposed EPReg achieved the best DSC, HD, and ASSD values on 41, 32, 38 sub-regions, respectively.

The numerical results on IXI30 are illustrated in **Figure 9**. IXI30 dataset has 95 subregions but 30 of them are extremely small regions. Thus we calculated DSC, HD, and ASSD values of the remaining 65 subregions. It can be observed that the proposed EPReg achieved the best DSC, HD, and ASSD values on 49, 35, 49 sub-regions, respectively, which shows that our network has satisfactory generalization ability. **Figure 10** visualizes registered images from different methods. Our network again attained overall satisfactory registration performance on this dataset.

3.5. Statistical Analyses

To investigate the statistical significance of the proposed network over other compared registration methods, a student test was conducted. Specifically, the two-sample, two-tailed t-test was employed to pairwisely compare the registration performance between our method and the other five methods on three different datasets (see **Table 3**). It can be observed that the null hypotheses for the five comparing pairs on the metric of DSC were not accepted at the 0.05 level. As a result, our method can be regarded as significantly better than the other five compared methods on DSC metric. In addition, the null hypotheses for the pairs of SyN-EPReg, VM-EPReg, and FAIM-EPReg on all three metrics were not accepted at the 0.05 level, which demonstrates our method was significantly better than these three methods on all three metrics. The p-values of PReg-EPReg on metrics of HD and ASSD from dataset Mindboggle101 were beyond the 0.05 level, which indicates that our method and PReg achieved similar performance with regard to the HD and ASSD evaluation on dataset Mindboggle101. The pair of MSNet-EPReg held the similar results on metrics of HD and ASSD from dataset IXI. In general, the statistical analyses prove that our method had an overall better registration performance than other compared cutting-edge methods.

3.6. Comparison of Time Efficiency

We further compared the time efficiency. **Table 4** lists the inference time for registering a pair of images using different methods. It can be observed that for the affine alignment, the time spent by our affine component (less than 0.3 s) was much less than that of the traditional affine alignment method (i.e., ANTS, Avants et al., 2008) using iterative optimization. Considering the overall registration time, our end-to-end registration network was much faster than other registration networks which require extra affine alignment (i.e., alignment using ANTS, Avants et al., 2008) before deformable registration.

4. DISCUSSION

Deformable registration is an important task of medical image computing and has various clinical applications. It is to search for the inhomogeneous point-wise displacements to match homologous locations from the moving domain to the fixed domain. Due to the large search space for the complicated non-rigid deformation, most existing registration schemes require separate rigid alignment before deformable registration to reduce the search space, or iteratively optimize the estimated deformation field. Even so, they may not well handle the large deformation cases. We have attempted to tackle this issue by devising a dual-stream deformable pyramid architecture.

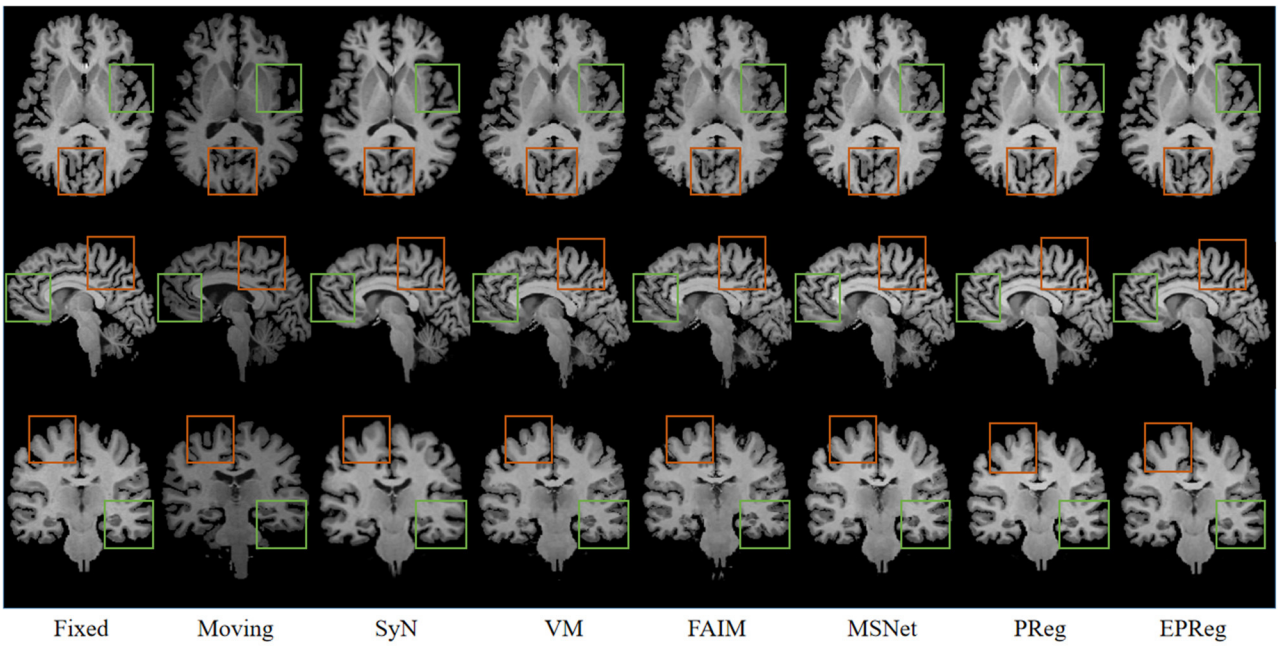


FIGURE 6 | Visualized registration results from different methods on Mindboggle101 dataset. Green and orange boxes highlight our accurate performance.

TABLE 1 | The DSC (%), Hausdorff distance (HD), and average symmetric surface distance (ASSD) results (mean ± SD) from SyN (Avants et al., 2008), VM (Balakrishnan et al., 2018), FAIM (Kuang and Schmah, 2019), MSNet (Duan et al., 2019), and our network on MindBoggle101 dataset.

	Methods	Frontal	Parietal	Occipital	Temporal	Cingulate
DSC (%)	SyN	54.4 ± 4.5	46.8 ± 6.8	49.8 ± 5.0	48.1 ± 6.2	49.2 ± 9.0
	VM	53.4 ± 8.1	52.7 ± 6.2	51.0 ± 7.3	43.3 ± 7.6	48.3 ± 9.2
	FAIM	57.2 ± 6.8	55.1 ± 7.1	53.7 ± 6.5	46.9 ± 6.9	50.8 ± 9.5
	MSNet	58.3 ± 7.1	50.4 ± 7.3	55.4 ± 6.9	47.7 ± 7.9	54.3 ± 8.9
	PReg	65.6 ± 9.3	60.5 ± 8.1	59.9 ± 7.5	67.5 ± 7.0	62.4 ± 11.2
	EPReg	**67.1 ± 9.0**	**61.5 ± 7.5**	**60.6 ± 7.0**	**67.9 ± 7.2**	**64.9 ± 9.5**
HD	SyN	13.2 ± 2.2	14.0 ± 1.9	14.8 ± 3.8	**6.8 ± 2.3**	7.9 ± 2.0
	VM	13.2 ± 3.6	13.6 ± 1.8	13.5 ± 4.6	8.6 ± 3.0	7.9 ± 1.9
	FAIM	12.9 ± 2.7	13.3 ± 1.7	13.3 ± 4.4	8.7 ± 2.3	7.8 ± 2.1
	MSNet	12.6 ± 2.9	13.2 ± 2.0	13.7 ± 4.4	8.0 ± 2.4	8.0 ± 2.2
	PReg	12.7 ± 3.2	13.5 ± 2.8	13.1 ± 3.7	7.4 ± 2.5	8.4 ± 2.1
	EPReg	**12.3 ± 3.1**	**13.2 ± 2.5**	**12.9 ± 3.6**	7.9 ± 2.3	**7.8 ± 2.0**
ASSD	SyN	1.32 ± 0.53	1.45 ± 0.32	1.48 ± 0.35	1.11 ± 0.27	1.37 ± 0.66
	VM	1.51 ± 0.33	1.32 ± 0.33	1.36 ± 0.32	1.32 ± 0.72	1.43 ± 0.43
	FAIM	1.42 ± 0.32	1.29 ± 0.28	1.44 ± 0.66	1.12 ± 0.46	1.20 ± 0.58
	MSNet	1.34 ± 0.28	1.29 ± 0.23	1.39 ± 0.53	1.06 ± 0.53	1.20 ± 0.63
	PReg	**0.42 ± 0.13**	**0.73 ± 0.22**	0.96 ± 0.32	0.62 ± 0.21	**1.00 ± 0.37**
	EPReg	0.43 ± 0.14	0.75 ± 0.26	**0.93 ± 0.30**	**0.61 ± 0.21**	1.03 ± 0.39

Best results are highlighted in bold.

The proposed deformable pyramid leverages multi-scale paired features to progressively estimate residual deformation field with a reduced search space rather than a large one, which facilitates the estimation of large/complicated deformation field. In addition, the affine alignment is also integrated within the deformable pyramid in a seamless manner, thus enables the trained network performing non-rigid registration in one forward pass.

Most unsupervised registration networks optimize deformation fields based on the maximization of the intensity-based image similarity. Considering the boundary/shape information is commonly used to constrain the registration,

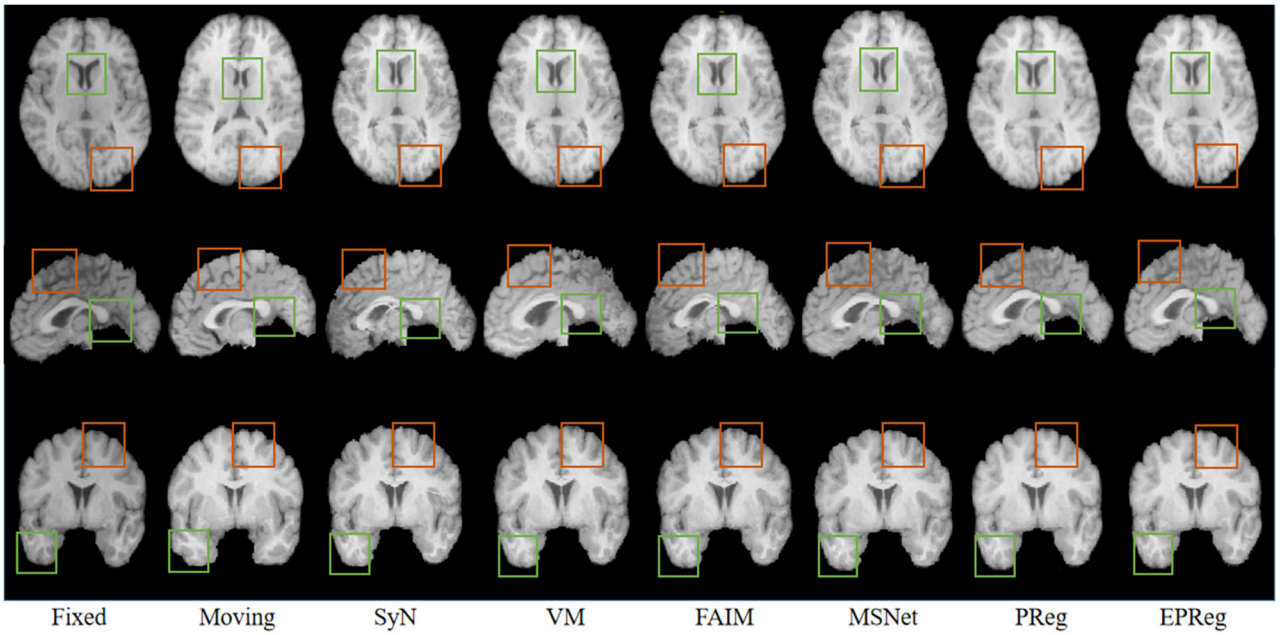


FIGURE 7 | Visualized registration results from different methods on LPBA40 dataset. Green and orange boxes highlight our accurate performance.

TABLE 2 | The DSC (%), Hausdorff distance (HD), and average symmetric surface distance (ASSD) results (mean ± SD) from SyN (Avants et al., 2008), VM (Balakrishnan et al., 2018), FAIM (Kuang and Schmah, 2019), MSNet (Duan et al., 2019), and our network on LPBA40 dataset.

	Methods	Frontal	Parietal	Occipital	Temporal	Cingulate	Putamen	Hippocampus
DSC (%)	SyN	68.1 ± 4.0	57.1 ± 5.8	50.7 ± 6.8	58.2 ± 4.9	58.9 ± 6.2	64.2 ± 7.3	65.2 ± 7.0
	VM	71.2 ± 3.1	57.4 ± 7.0	58.8 ± 6.0	61.7 ± 4.4	54.9 ± 6.2	66.1 ± 9.4	60.9 ± 9.0
	FAIM	73.1 ± 3.1	59.4 ± 7.3	60.6 ± 6.1	65.1 ± 4.2	61.2 ± 5.8	70.8 ± 10.0	68.0 ± 6.4
	MSNet	73.4 ± 3.3	60.0 ± 7.0	61.0 ± 6.1	65.9 ± 4.0	62.1 ± 5.5	72.7 ± 8.5	68.8 ± 5.7
	PReg	76.8 ± 2.7	66.6 ± 6.2	66.7 ± 4.8	71.6 ± 3.4	65.9 ± 6.6	78.6 ± 2.6	75.2 ± 2.7
	EPReg	**77.0 ± 2.6**	**68.3 ± 4.5**	**68.7 ± 3.7**	**72.7 ± 2.7**	**66.7 ± 5.4**	**79.1 ± 2.5**	**75.3 ± 2.8**
HD	SyN	13.2 ± 1.8	**15.8 ± 2.5**	16.2 ± 2.9	16.9 ± 3.0	13.0 ± 3.1	10.9 ± 9.9	7.1 ± 1.4
	VM	13.6 ± 1.9	19.2 ± 5.0	15.1 ± 2.9	16.8 ± 3.4	13.6 ± 2.8	8.9 ± 2.7	8.5 ± 2.0
	FAIM	13.5 ± 2.0	19.1 ± 4.8	15.1 ± 3.0	16.5 ± 3.4	13.2 ± 2.6	8.4 ± 2.9	7.9 ± 1.9
	MSNet	13.5 ± 1.9	19.0 ± 4.8	15.0 ± 2.9	16.4 ± 3.4	13.1 ± 2.7	8.2 ± 3.1	7.6 ± 1.8
	PReg	12.3 ± 1.5	18.1 ± 4.8	14.6 ± 2.9	15.8 ± 3.4	13.2 ± 2.7	7.6 ± 3.1	6.4 ± 1.8
	EPReg	**12.3 ± 1.4**	17.5 ± 4.0	**14.1 ± 2.3**	**15.0 ± 2.7**	**12.5 ± 2.1**	**7.4 ± 3.2**	**6.2 ± 1.3**
ASSD	SyN	2.44 ± 0.44	2.71 ± 0.40	3.18 ± 0.58	2.55 ± 0.40	2.22 ± 0.31	1.70 ± 0.63	1.56 ± 0.33
	VM	1.41 ± 0.27	2.11 ± 0.41	1.87 ± 0.30	1.49 ± 0.19	2.02 ± 0.31	1.28 ± 0.32	1.41 ± 0.32
	FAIM	1.35 ± 0.27	2.06 ± 0.41	1.86 ± 0.32	1.38 ± 0.18	1.87 ± 0.29	1.12 ± 0.32	1.22 ± 0.23
	MSNet	1.31 ± 0.26	2.04 ± 0.41	1.86 ± 0.31	1.35 ± 0.18	1.82 ± 0.27	1.05 ± 0.28	1.18 ± 0.21
	PReg	1.11 ± 0.23	1.72 ± 0.37	1.60 ± 0.21	1.16 ± 0.15	1.68 ± 0.31	0.85 ± 0.17	0.98 ± 0.11
	EPReg	**1.08 ± 0.19**	**1.62 ± 0.26**	**1.54 ± 0.18**	**1.11 ± 0.11**	**1.65 ± 0.25**	**0.85 ± 0.16**	**0.96 ± 0.11**

weakly supervised registration networks pay extra attention on leveraging the correspondences between structural information. However, obtaining such corresponding structures requires manual annotations. Instead of utilizing the structural information to form an explicit loss function, we simply apply the Sobel edge map as an extra input of the network. The purpose is to enhance the texture structures of image content, thus to pay attention for the edge-aware alignment. The comparison results listed in **Tables 1**, **2** (evaluations on some relatively larger regions) show that the network with edge-aware input (i.e., EPReg) achieved overall better registration performance than PReg. The comparisons on lots of small sub-regions shown in **Figures 8**, **9** demonstrate that EPReg attained more



FIGURE 8 | Comparisons of DSC (%), HD, and ASSD results by different methods on LPBA40 dataset. The results were evaluated across the 54 corresponding ROIs in LPBA40 dataset, "*" indicates that the proposed EPReg outperformed other methods.

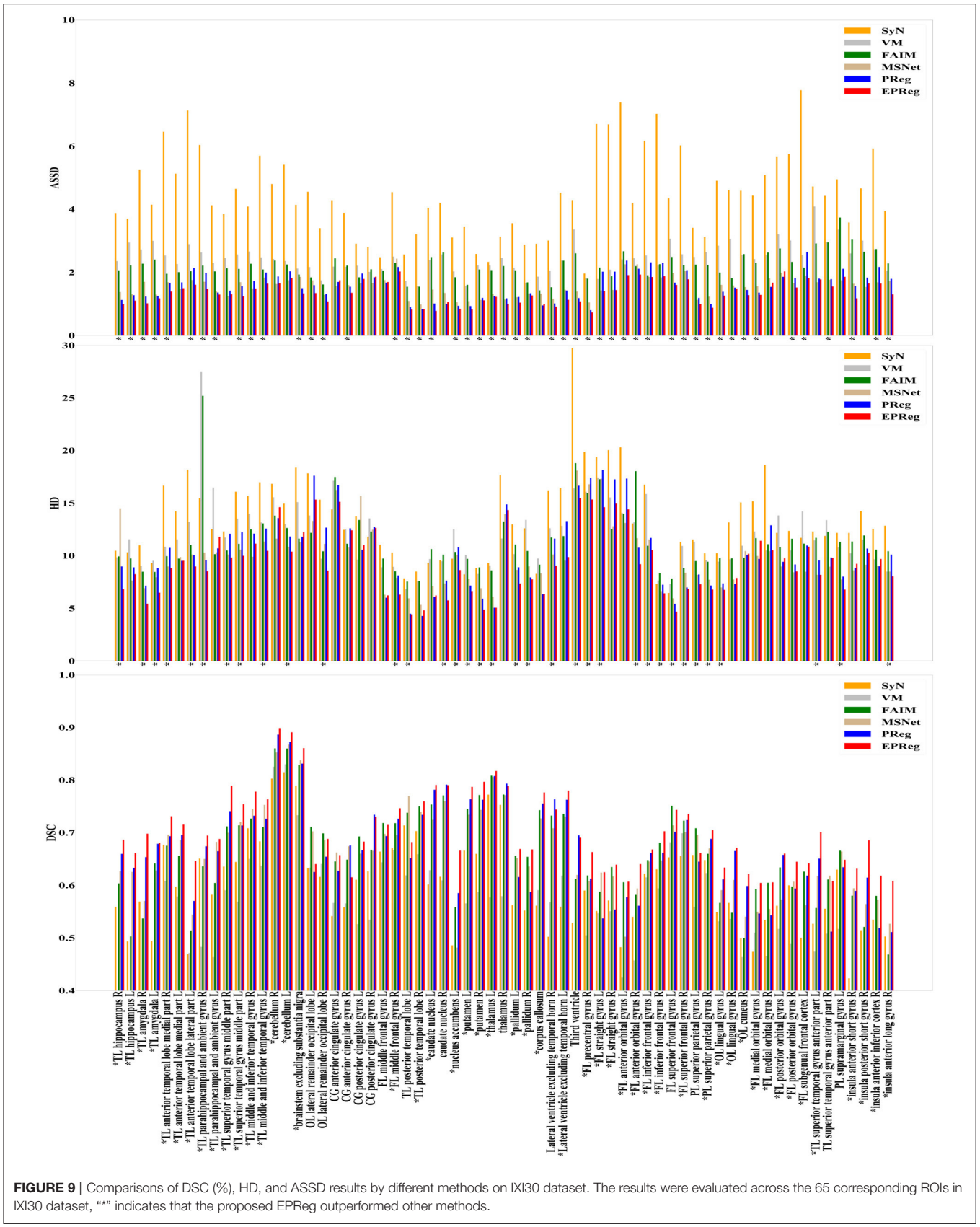


FIGURE 9 | Comparisons of DSC (%), HD, and ASSD results by different methods on IXI30 dataset. The results were evaluated across the 65 corresponding ROIs in IXI30 dataset, "*" indicates that the proposed EPReg outperformed other methods.

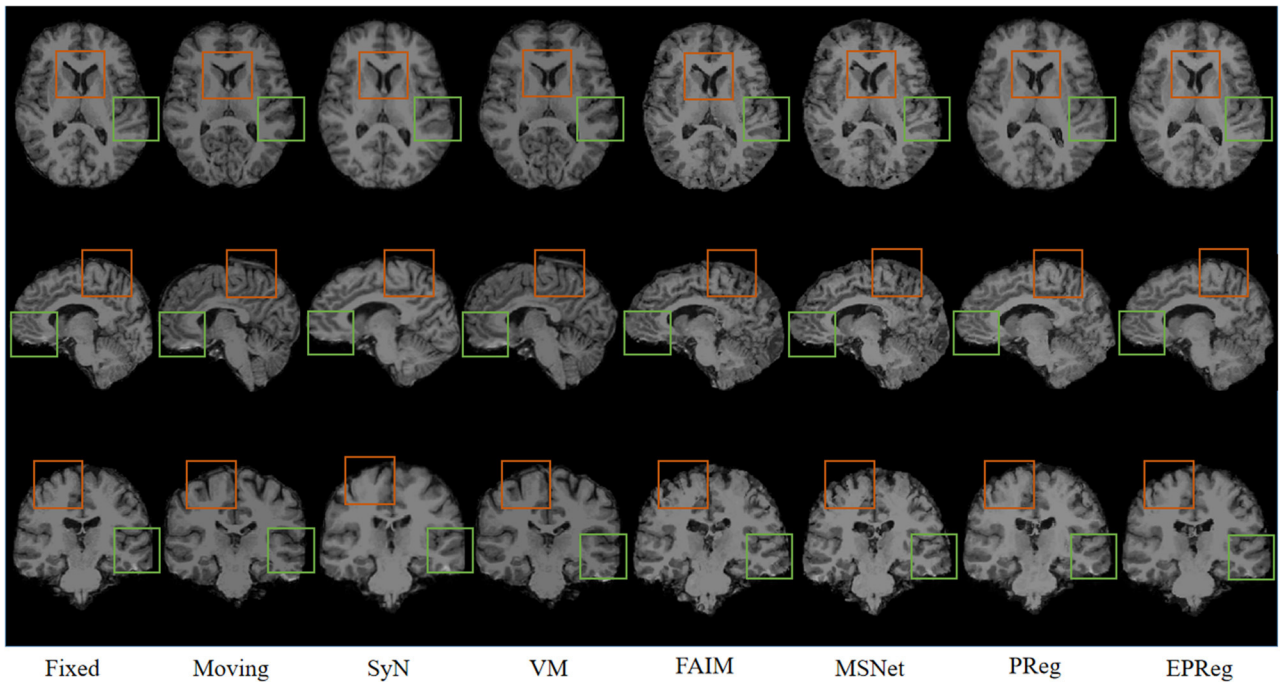


FIGURE 10 | Visualized registration results from different methods on IXI30 dataset. Green and orange boxes highlight our accurate performance.

TABLE 3 | *P*-values of student tests between different methods on different metrics.

Method	Mindboggle101			LPBA40			IXI		
	DSC	HD	ASSD	DSC	HD	ASSD	DSC	HD	ASSD
SyN-EPReg	8.35e-13	4.01e-02	9.02e-06	3.84e-05	2.13e-07	4.79e-03	1.30e-03	3.15e-02	7.60e-03
VM-EPReg	1.47e-14	2.45e-03	9.80e-09	6.93e-16	7.16e-16	7.88e-09	1.11e-04	2.16e-02	1.17e-04
FAIM-EPReg	9.70e-17	1.72e-03	5.32e-08	1.22e-14	5.73e-16	2.50e-08	6.09e-05	4.54e-02	6.72e-03
MSNet-EPReg	5.93e-12	2.08e-03	6.91e-08	3.56e-10	2.07e-16	1.58e-10	1.56e-03	0.91	0.71
PReg-EPReg	1.84e-06	0.63	0.65	2.00e-05	3.74e-05	2.40e-05	4.50e-03	6.89e-03	2.49e-02

TABLE 4 | Inference time (*second*) for registering a pair of images using different methods.

Methods	Mindboggle101	LPBA40	IXI
ANTS (affine alignment)	8.19 ± 0.30	7.73 ± 0.27	9.13 ± 0.29
SyN	39.24 ± 2.07	33.32 ± 1.46	40.10 ± 2.01
VM	0.66 ± 0.01	0.11 ± 0.01	0.73 ± 0.01
FAIM	1.17 ± 0.01	0.49 ± 0.01	1.26 ± 0.01
MSNet	1.34 ± 0.01	0.45 ± 0.01	1.41 ± 0.01
EPReg (affine only)	0.09 ± 0.01	0.09 ± 0.01	0.23 ± 0.01
EPReg	0.25 ± 0.01	0.23 ± 0.01	0.61 ± 0.01

accurate registration than PReg, due to the usage of the edge-aware information.

We evaluated the proposed edge-aware pyramidal deformable network on three different brain MR datasets and compared with four cutting-edge registration methods. The DSC, HD, and ASSD were employed to evaluate all methods. Specifically, HD and ASSD were employed to provide evaluation on the differences in boundary shape between two volumes. The numerical results and statistical analyses show that our method had an overall better registration accuracy than other compared cutting-edge methods. Furthermore, our registration network provided very efficient inference procedure,

which achieved accurate volumetric registration within 1 s. From the experimental results, it can be demonstrated that the proposed progressive deformation estimation scheme contributed to the improvement of registration accuracy and efficiency.

The proposed registration method focuses on the large deformation estimation by progressively predicting the residual deformation. It is beneficial for the estimation of large-scale deformation. Further validations on other body parts (e.g., chest or abdominal CT images) will be our future work.

The problem of registration validation in clinical settings is still an open issue. Most of recent research articles focusing on developing new registration approaches employ DSC as a primary metric to evaluate the registration accuracy (Balakrishnan et al., 2018; Cao et al., 2018; Loi et al., 2018, 2020; Fan et al., 2019; Huang et al., 2021). For a fair comparison, we also applied DSC to evaluate the registration performance. However, the DSC may suffer of the limitation to be dependent on the volume of structures. Thus we further employed HD and ASSD to evaluate boundary differences between two volumetric regions. Although there is no guaranteed thresholds w.r.t. DSC, HD, ASSD for quality assessment of registration on brain MR images, the comparison results on a series of sub-regions (over 50 sub-regions) show the proposed network consistently outperformed other cutting-edge registration methods with respect to metrics of DSC, HD, and ASSD. In addition, the visual results in **Figures 6**, **7**, **10** illustrate the satisfactory registration performance obtained by our method. In our future work, the phantom study which could generate the ground truth deformation fields for straightforward registration validation would be conducted.

5. CONCLUSION

In this paper, we have presented an edge-aware pyramidal deformable network for unsupervised volumetric registration. The proposed network focuses on the large deformation estimation by progressively predicting the residual deformation. Our key idea is to fully exploit the useful complementary multi-level information from paired features to predict multi-scale deformation fields. We achieve this by developing a deformable pyramid architecture, which can generate multi-scale paired feature maps for progressively transforming the paired information into more representative features to predict more accurate deformation field. In addition, we leverage extra edge information to impel the network pay attention to the edge-aware alignment. Extensive experiments on several 3D MRI datasets demonstrate that our edge-aware pyramidal deformable network achieves satisfactory registration performance. The coarse-to-fine progressive registration procedure is beneficial to compensate for the large-scale deformation, and can be regarded as a general solution for deformable volumetric registration.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YC, ZZ, YR, CQ, DL, QD, DN, and YW response for study design. YC and CQ implemented the research. YC, ZZ, QD, and YW conceived the experiments. YC, ZZ, and YR conducted the experiments. YC, ZZ, YR, DL, and YW analyzed the results. YC and YW wrote the main manuscript text and prepared the figures. All authors reviewed the manuscript.

FUNDING

This work was supported in part by the National Key R&D Program of China (No. 2019YFC0118300), in part by the National Natural Science Foundation of China under Grants 62071305 and 61701312, in part by the Guangdong Basic and Applied Basic Research Foundation (2019A1515010847), in part by the Medical Science and Technology Foundation of Guangdong Province (B2019046), in part by the Natural Science Foundation of SZU (No. 860-000002110129), and in part by the Shenzhen Peacock Plan (KQTD2016053112051497).

REFERENCES

Auzias, G., Colliot, O., Glaunes, J. A., Perrot, M., Mangin, J.-F., Trouve, A., et al. (2011). Diffeomorphic brain registration under exhaustive sulcal constraints. *IEEE Trans. Med. Imaging* 30, 1214–1227. doi: 10.1109/TMI.2011.2108665

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

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2018). "An unsupervised learning model for deformable medical image registration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 9252–9260.

Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision* 61, 139–157. doi: 10.1023/B:VISI.0000043755.93987.aa

Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., et al. (2017). "Deformable image registration based on similarity-steered cnn regression," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 300–308.

Cao, X., Yang, J., Zhang, J., Wang, Q., Yap, P. T., and Shen, D. (2018). Deformable image registration using a cue-aware deep regression network. *IEEE Trans. Biomed. Eng.* 65, 1900–1911. doi: 10.1109/TBME.2018.2822826

de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* 52, 128–143. doi: 10.1016/j.media.2018.11.010

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302.

Duan, L., Yuan, G., Gong, L., Fu, T., Yang, X., Chen, X., et al. (2019). Adversarial learning for deformable registration of brain mr image using

a multi-scale fully convolutional network. *Biomed. Signal Process. Control* 53:101562. doi: 10.1016/j.bspc.2019.101562

Eppenhof, K. A., Lafarge, M. W., Moeskops, P., Veta, M., and Pluim, J. P. (2018). "Deformable image registration using convolutional neural networks," in *Medical Imaging 2018: Image Processing*, Vol. 10574 (San Francisco, CA: International Society for Optics and Photonics).

Fan, J., Cao, X., Yap, P.-T., and Shen, D. (2019). Birnet: brain image registration using dual-supervised fully convolutional networks. *Med. Image Anal.* 54, 193–206. doi: 10.1016/j.media.2019.03.006

Fischl, B. (2012). Freesurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Haskins, G., Kruger, U., and Yan, P. (2020). Deep learning in medical image registration: a survey. *Mach. Vis. Appl.* 31:8. doi: 10.1007/s00138-020-01060-x

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: IEEE Computer Society), 770–778.

Hu, X., Kang, M., Huang, W., Scott, M. R., Wiest, R., and Reyes, M. (2019). "Dual-stream pyramid registration network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 382–390.

Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., et al. (2018). Weakly-supervised convolutional neural networks for multimodal image registration. *Med. Image Anal.* 49, 1–13. doi: 10.1016/j.media.2018.07.002

Huang, Y., Ahmad, S., Fan, J., Shen, D., and Yap, P.-T. (2021). Difficulty-aware hierarchical convolutional neural networks for deformable registration of brain mr images. *Med. Image Anal.* 67:101817. doi: 10.1016/j.media.2020.101817

Huttenlocher, D. P., Klanderman, G. A., and Rucklidge, W. J. (1993). Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 850–863. doi: 10.1109/34.232073

Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. (2015). "Spatial transformer networks," in *Advances in Neural Information Processing Systems* (Montreal, QC), 2017–2025.

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint* arXiv:1412.6980.

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46, 786–802. doi: 10.1016/j.neuroimage.2008.12.037

Klein, A., and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* 6:171. doi: 10.3389/fnins.2012.00171

Knops, Z. F., Maintz, J. A., Viergever, M. A., and Pluim, J. P. (2006). Normalized mutual information based registration using k-means clustering and shading correction. *Med. Image Anal.* 10, 432–439. doi: 10.1016/j.media.2005.03.009

Kuang, D., and Schmah, T. (2019). "Faim–a convnet method for unsupervised 3d medical image registration," in *International Workshop on Machine Learning in Medical Imaging* (Shenzhen: Springer), 646–654.

Li, H., and Fan, Y. (2018). "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC: IEEE), 1075–1078.

Loi, G., Fusella, M., Lanzi, E., Cagni, E., Garibaldi, C., Iacoviello, G., et al. (2018). Performance of commercially available deformable image registration platforms for contour propagation using patient-based computational phantoms: a multi-institutional study. *Med. Phys.* 45, 748–757. doi: 10.1002/mp.12737

Loi, G., Fusella, M., Vecchi, C., Menna, S., Rosica, F., Gino, E., et al. (2020). Computed tomography to cone beam computed tomography deformable image registration for contour propagation using head and neck, patient-based computational phantoms: a multicenter study. *Pract. Radiat. Oncol.* 10, 125–132. doi: 10.1016/j.prro.2019.11.011

Rao, Y. R., Prathapani, N., and Nagabhooshanam, E. (2014). Application of normalized cross correlation to image registration. *Int. J. Res. Eng. Technol.* 3, 12–16. doi: 10.15623/ijret.2014.0317003

Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., and Pennec, X. (2017). "Svf-net: Learning deformable image registration using shape matching," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 266–274.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 234–241.

Serag, A., Aljabar, P., Ball, G., Counsell, S. J., Boardman, J. P., Rutherford, M. A., et al. (2012). Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *Neuroimage* 59, 2255–2265. doi: 10.1016/j.neuroimage.2011.09.062

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3d probabilistic atlas of human cortical structures. *Neuroimage* 39, 1064–1080. doi: 10.1016/j.neuroimage.2007.09.031

Sheikhjafari, A., Noga, M., Punithakumar, K., and Ray, N. (2018). Unsupervised deformable image registration with fully connected generative neural network. in *1st Conference on Medical Imaging with Deep Learning (MIDL 2018)* (Amsterdam), 1–8.

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

Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B. P., Išgum, I., and Staring, M. (2017). "Nonrigid image registration using multi-scale 3d convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 232–239.

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

Taha, A. A., and Hanbury, A. (2015). Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging* 15:29. doi: 10.1186/s12880-015-0068-x

Uzunova, H., Wilms, M., Handels, H., and Ehrhardt, J. (2017). "Training cnns for image registration from few samples with model-based data augmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Quebec City, QC: Springer), 223–231.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage* 45, S61–S72. doi: 10.1016/j.neuroimage.2008.10.040

Wang, Y., Dou, H., Hu, X., Zhu, L., Yang, X., Xu, M., et al. (2019). Deep attentive features for prostate segmentation in 3d transrectal ultrasound. *IEEE Trans. Med. Imaging* 38, 2768–2778. doi: 10.1109/TMI.2019.2913184

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Wolberg, G., and Zokai, S. (2000). "Robust image registration using log-polar transform," in *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, Vol. 1 (Vancouver, BC: IEEE), 493–496.

Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017). Quicksilver: fast predictive image registration–a deep learning approach. *Neuroimage* 158, 378–396. doi: 10.1016/j.neuroimage.2017.07.008

Zhao, S., Dong, Y., Chang, E. I., Xu, Y., et al. (2019a). "Recursive cascaded networks for unsupervised medical image registration," in *Proceedings of the IEEE International Conference on Computer Vision* (Seoul), 10600–10610.

Zhao, S., Lau, T., Luo, J., Eric, I., Chang, C., and Xu, Y. (2019b). Unsupervised 3d end-to-end medical image registration with volume tweening network. *IEEE J. Biomed. Health Inform.* 24, 1394–1404. doi: 10.1109/JBHI.2019.2951024

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 January 2021
doi: 10.3389/fnins.2020.629572

An Inter- and Intra-Subject Transfer Calibration Scheme for Improving Feedback Performance of Sensorimotor Rhythm-Based BCI Rehabilitation

Lei Cao[1], *Shugeng Chen*[2], *Jie Jia*[2*], *Chunjiang Fan*[3*], *Haoran Wang*[4] *and Zhixiong Xu*[1]

[1] Department of Artificial Intelligence, Shanghai Maritime University, Shanghai, China, [2] Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai, China, [3] Wuxi Rehabilitation Hospital, Wuxi, China, [4] Department of Computer Science and Technology, Tongji University, Shanghai, China

The Brain Computer Interface (BCI) system is a typical neurophysiological application which helps paralyzed patients with human-machine communication. Stroke patients with motor disabilities are able to perform BCI tasks for clinical rehabilitation. This paper proposes an effective scheme of transfer calibration for BCI rehabilitation. The inter- and intra-subject transfer learning approaches can improve the low-precision classification performance for experimental feedback. The results imply that the systematical scheme is positive in increasing the confidence of voluntary training for stroke patients. In addition, it also reduces the time consumption of classifier calibration.

Keywords: BCI, transfer learning, rehabilitation, stroke, classifier calibration

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Shangce Gao,
University of Toyama, Japan
Cheng Wang,
Huaqiao University, China

***Correspondence:**
Chunjiang Fan
lcao@shmtu.edu.cn
Jie Jia
shannonjj@126.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *15 November 2020*
Accepted: *21 December 2020*
Published: *28 January 2021*

Citation:
Cao L, Chen S, Jia J, Fan C, Wang H and Xu Z (2021) An Inter- and Intra-Subject Transfer Calibration Scheme for Improving Feedback Performance of Sensorimotor Rhythm-Based BCI Rehabilitation. Front. Neurosci. 14:629572. doi: 10.3389/fnins.2020.629572

1. INTRODUCTION

Brain-computer interface (BCI) is developed as an extrinsic pathway for human-machine interaction in a reliable way (Birbaumer, 2006). It is effective for the disabled to control external devices by neural activities (Buch et al., 2008). Stroke patients with motor disabilities in particular, are able to perform BCI tasks for clinical rehabilitation (Meng et al., 2016). In this treatment, sensorimotor rhythm changes are used as neurological modulation for active intervention (Mane et al., 2019).

During rehabilitation, the patients are requested to attempt or to imagine performing a movement. Then, motor attempt (MA) or motor imagery (MI)-BCI systems will output a synchronized sensory biofeedback (e.g., robotic arm recovery) by a trained classifier based on a prior dataset (Pillette et al., 2020). In the intervention, the functional motor is significantly enabled by neurophysiological activity (Xu et al., 2014). This is an ongoing process of brain plasticity and functional recovery (Remsik et al., 2019). Recent studies have reported the improvement of limb movement for stroke patients using long-term sensorimotor rhythm (SMR)-BCI interventions (Ramos-Murguialday et al., 2013; Pichiorri et al., 2015; Bundy et al., 2017).

Nevertheless, BCI rehabilitation is limited by poor-efficiency recognition algorithms and model-personalized variability (Grosse-Wentrup et al., 2011). Relevant work has proved that BCI decoding accuracy was insufficient for rehabilitation outcomes (Mane et al., 2020). Moreover, the failures of BCI feedback also reduce the confidence of trainees subtly (Foong et al., 2019). Hence, various improvements of pattern recognition and model calibration should be made to enhance SMR-BCI performance in the clinical application.

Conventionally, SMR features are effectively extracted by a time-frequency analysis for the healthy (Pfurtscheller and Da Silva, 1999). For instance, the event-related desynchronization (ERD) amplitudes oscillated in the μ rhythm was detected during the motor imagery task for pattern recognition (Huang et al., 2012; Saha et al., 2017). Furthermore, the common spatial pattern (CSP) algorithm was proposed for feature extracting in the spatial domain (Wang et al., 2006; Arvaneh et al., 2013). It was efficient for mining the significant difference of two-type motor tasks. However, the degeneration of neural activation due to post stroke, has a negative impact on BCI performance (Shu et al., 2018). Thus, the signal characteristics are much lower than those of healthy individuals during motor tasks (De Vries et al., 2013; Caria et al., 2020). Therefore, increasing the precision of SMR-BCI using a mathematical methods is meaningful for BCI intervention.

To solve this problem, transfer learning (TL), which applies the dataset in source domains for compensating insufficient labeled data in a target domain, has been proposed for MI-BCIs (Samek et al., 2013b; Azab et al., 2018). This technology is developed in several ways, such as instance selection (Wu, 2016; Hossain et al., 2018), feature calibration (Samek et al., 2013a; Zhao et al., 2019) and classification domains (Vidaurre et al., 2010; He and Wu, 2019). For instance, for selection, active learning is typically presented for selecting training data from intra- or inter-subject labeled trials (Hossain et al., 2018). The target of this approach is to increase the informative trials of the new subject by adding sufficient existing labeled trials that were close to prior dataset. In the feature domain, transfer calibration approaches mainly concentrate on regulating the covariance matrix estimation and optimization function for improving the performance of CSP models. For example, researchers regularized the CSP filter by the average of the common feature space from other subjects (Kang et al., 2009). Moreover, the efficiency of domain adaptation has been verified for MI-BCI in the classification domain (Vidaurre et al., 2010). Kobler et al. constructed a Restricted Boltzmann Machine (RBM) based on public baseline data and applied it for the MI-BCI task (Kobler and Scherer, 2016). Recently, multi-task learning has been presented in the relevant experiment (Jayaram et al., 2016; Gao et al., 2019), where the weight parameters of inter-subjective classifiers were learned jointly for minimizing the dissimilarities between these existing classification models and the target model. However, BCIs controlled by these approaches have been proven to be only valid for healthy individuals. None of them are experimentally evaluated for BCI rehabilitation.

This paper first proposes a transfer calibration scheme to improve the rehabilitation outcomes of SMR-BCI. First, we utilize a transfer learning algorithm, whose effects have been verified in the task of MI-BCI (Azab et al., 2019), to validate the reliability for stroke patients. Then, we discuss the respective applicability between intra- and inter- subjective conditions. The results show that our proposed approaches improved the low-precision classification performance. Accordingly, we generalize the scheme of transfer calibration for SMR-BCI intervention. This methodology applies for other transfer learning algorithms to increase the precision of BCI feedback.

2. MATERIALS AND METHODS

2.1. Experimental Paradigm and Subjects

Seven stroke patients aged between 30 and 65, recruited from the Department of Rehabilitation Medicine of Huashan Hospital participated in our experiments (**Table 1**). All of them were naive to BCI and provided consent to be involved in the study. The inclusion criteria for this study were as follows: (1) unilateral motor dysfunction diagnosed by computer tomography or magnetic resonance imaging (MRI); (2) first onset stroke patient; (3) the time since stroke onset was more than 4 weeks and less than 6 months; (4) the assessment of cognitive functions: Mini-Mental State Examination score $>$ 25. The exclusion criteria were listed below: (1) unstable medical conditions; (2) severe vision problems; (3) the intervention treatment by other brain stimulations during the study period.

In the experiment, BCI intervention was performed in three sessions a week for each patient. And it lasted 1 month, with a total of 12 sessions. One session contained three runs, each run had 30 trials for each mental task (motor attempt or idle state) performed in a random order. Subjects underwent two experimental tasks. In the task of motor attempt (MA), patients were required to attempt motion of wrist extension with affected hands continually, but not to have compensatory movements. In the other task of idle state (IS), they needed to do nothing but rest (**Figure 1A**).

The experimental paradigm is shown in **Figure 1B**. In one trial, the duration time was about 11 s. The patient was asked to sit in front of a computer screen, with arms resting on a desk. A white arrow presented on the center of the screen from 0 to 3 s. The patient was instructed to keep still and rest. Then, an alternative indicator (a red square or a red rectangle) was displayed in the center of the screen, representing a task (MA or IS). After the command disappeared, the subject was required to perform the corresponding task for 5 s until the white cross disappeared. Finally, the rest interval was adopted randomly to relax.

2.2. EEG Recording and Signal Preprocessing

EEG were collected using 32 channels consisting of Ag/AgCl electrode of EEG cap (actiCAP; Brain Products, Germany) according to the configuration of 10-C20 International System. A bio-signal amplifier (Brain Products) was used for acquiring the signals. The unilateral reference electrode was located in the right mastoid process, and the ground electrode was located in the forehead. The other 31 channels (FP1, FZ, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, PZ, P3, P7, O1, O2, P4, P8, TP10, CP6, CP2, CZ, C4, T8, FT10, FC6, FC2, F4, F8, FP2) were used for calculation. Electrode impedances were kept below 5 kΩ. The signals were amplified, digitalized with a sample rate of 200 Hz, and bandpass-filtered between 1 and 35 Hz.

2.3. Evaluation of BCI Performance

BCI performance was evaluated by classification accuracy for two mental tasks. As shown in the figure of experimental paradigm, the subject conducted the mental task from 4.5 s to 9.5 s. Hence,

TABLE 1 | The baseline clinical characteristics of participants.

Patients	Gender	Affected hand	Type of injury	Time since injury (m)	Site of injury
S1	Male	Right	Ischemia	5	Left, basal ganglia
S2	Male	Left	Hemorrhage	4	Right, basal ganglia
S3	Male	Right	Hemorrhage	1	Left, basal ganglia
S4	Male	Right	Ischemia	1	Left, paracele
S5	Male	Right	Ischemia	3	Left, basal ganglia
S6	Male	Right	Ischemia	5	Left, paracele, basal ganglia
S7	Male	Left	Ischemia	3	Right, basal ganglia

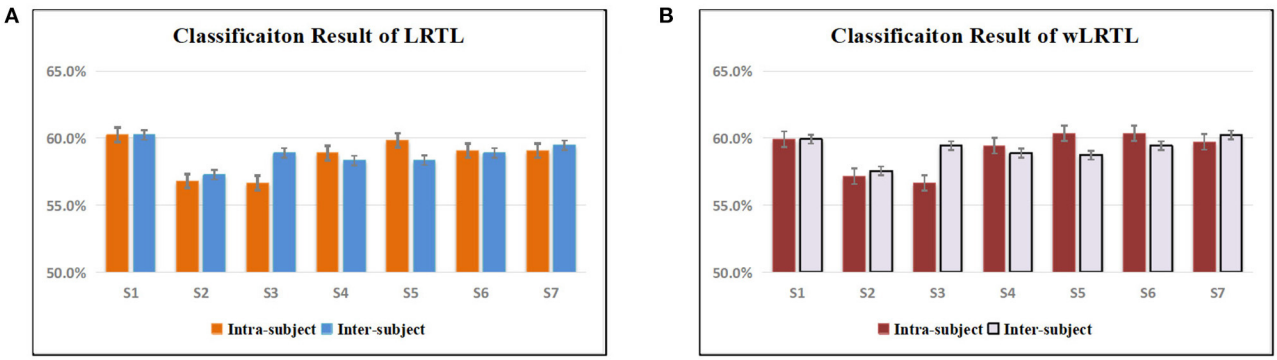


FIGURE 1 | (A) Experimental procedures of two different mental tasks. In the task of MA, patients were required to attempt motion of wrist extension with affected hands continually, but not to have compensatory movements. In the other task of IS, they need not do anything for a rest. **(B)** Overview of BCI control paradigm.

EEG signals were extracted from [4.5 9.5] s of each single trial. And the features were filtered with the common spatial pattern (CSP) method, whose log-variance of the first and last three components were selected as output vectors. Then, a transfer learning scheme was used to improve the classification efficiency. The Logistic Regression (LR) -based classifier was used as the baseline approach. Its classification parameters w_t were calibrated using the following function,

$$L(w_t) = \min_{w_t} \left(\sum_{n_t}^{i=1} H(w_t; l_t^i, f_t^i) + \lambda_t \|w_t\|_2^2 \right), \quad (1)$$

where H, f_t, l_t and $\|.\|_2$ denote the cross-entropy, feature vectors extracted from the EEG dataset, label vector, and 2-norm functions, respectively. Conventionally, the parameters were supposed be trained by large prior data. Transfer learning algorithms had been proposed for improving the efficiency of parameter calibration of new subjects without subject-specific data (Hossain et al., 2018). In the framework of logistic regression -based transfer learning (LRTL), a regularization term penalizing dissimilarities R_t was used for transferring the prior distribution of the existing classification parameters into the calibration of the present training of new target subjects or sessions. In this opinion, the classification parameters were calculated as follows,

$$L'(w_t) = \min_{w_t} \left(\sum_{n_t}^{i=1} H(w_t; l_i^t, l_i^f) + \lambda_t R_t(w_t) \right), \quad (2)$$

and the R_t was decided by estimating the similarity between feature distribution of existing models and that of current few training data,

$$R_t(w_t) = 0.5[(w_t - \mu)^T \Sigma_t^{-1}(w_t - \mu) + \log(|\Sigma_t|)], \quad (3)$$

where μ and Σ_t were, respectively, obtained as

$$\mu = \frac{1}{n} \sum_{t'=1}^{n} w_{t'}, \quad (4)$$

$$\Sigma_t = \frac{diag\left(\Sigma_{t'=1}^{n}(w_{t'} - \mu)(w_{t'} - \mu)^T\right)}{trace\left(\Sigma_{t'=1}^{n}(w_{t'} - \mu)(w_{t'} - \mu)^T\right)}. \quad (5)$$

Here, T denoted transpose of the matrix, the functions of *diag* and *trace* were defined as the diagonal elements and the sum of the diagonal elements of a matrix. Furthermore, Kullback-Leibler (KL) divergence was added to solve the problem of weight distribution between existing models and the target model. It was supposed to give larger weights to more similar distributions and smaller weights to less similar distributions. KL divergence of two EEG sets (E_0, E_1) were represented as the following form,

$$KL[E_0 \,||\, E_1] = 0.5[(\mu_1 - \mu_0)^T \Sigma_1^{-1}(\mu_1 - \mu_0) + trace(\Sigma_1^{-1}\Sigma_0) - ln\left(\frac{det(\Sigma_0)}{det(\Sigma_1)}\right) - K], \quad (6)$$

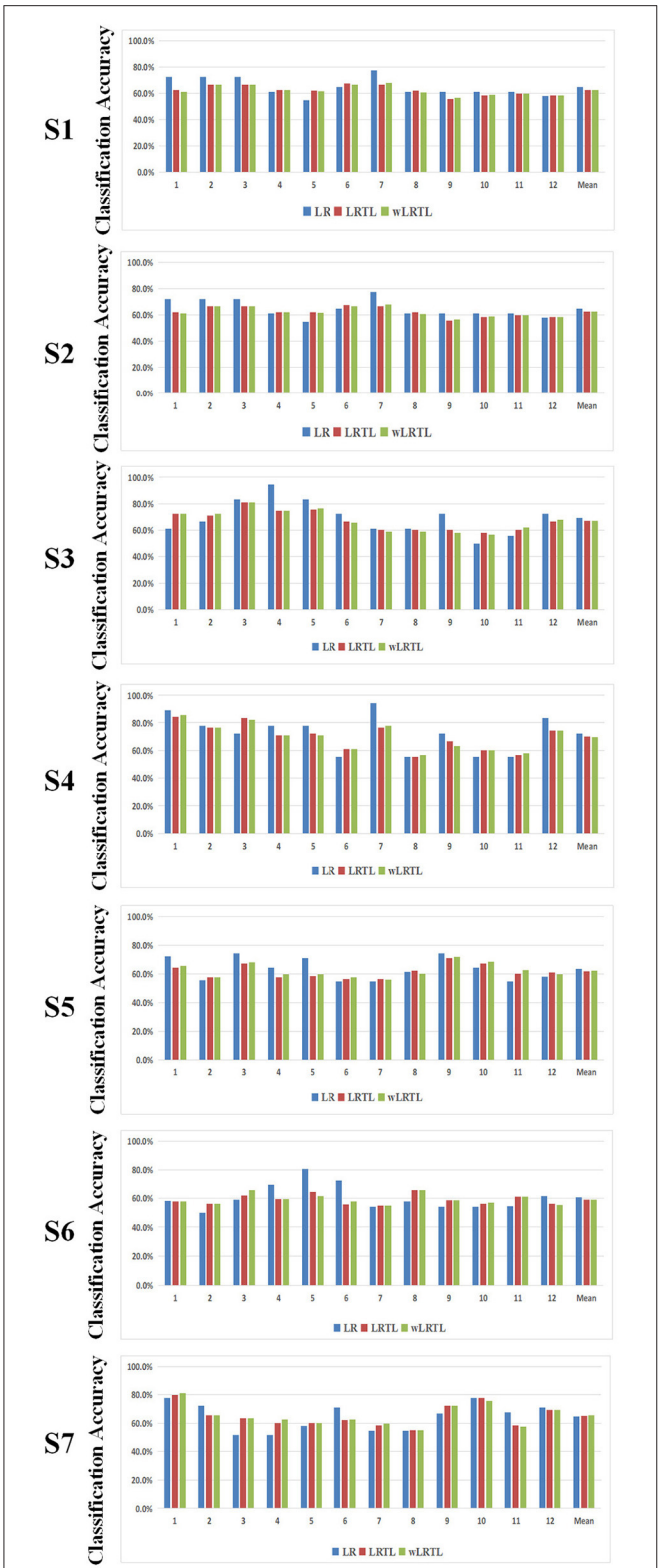


FIGURE 2 | The classification results of transfer learning algorithms and baseline algorithm under the intra-subject condition. Specially, the accuracies were missing for LRTL and wLRTL. Because the new subject had no experience of relevant tasks. Hence, the transfer calibration was not performed without prior dataset under this condition.

TABLE 2 | The average accuracy of low-precision sessions (≤ 60%) for transfer learning algorithms (i.e., LRTL, wLRTL) and baseline algorithm (i.e., LR) under the intra-subject condition.

Methods	S1	S2	S3	S4	S5	S6	S7
LR (%)	56.5	53.6	52.8	55.6	55.6	54.7	54.2
LRTL (%)	60.2	56.8	56.7	58.9	58.2	59.8	59.1
wLRTL (%)	59.9	57.2	56.7	59.4	58.6	60.4	59.7

The accuracies of transfer learning approaches were higher than that of baseline approach.

TABLE 3 | The average accuracy of low-precision sessions (≤ 60%) for transfer learning algorithms (i.e., LRTL, wLRTL) and baseline algorithm (i.e., LR) under the inter-subject condition.

Methods	S1	S2	S3	S4	S5	S6	S7
LR (%)	56.5	52.9	52.8	55.6	55.6	55.1	54.2
LRTL (%)	60.2	57.3	58.9	58.3	58.3	58.9	59.5
wLRTL (%)	59.9	57.6	59.4	58.9	58.7	59.4	60.2

The accuracies of transfer learning approaches were higher than that of baseline approach.

where *det* and K denoted the determinant function and the dimension of the data, respectively. Thus, the Σ_t was updated as below,

$$\Sigma_{t_w} = \frac{diag\left(\Sigma_{t'=1}^{n}(\alpha_{t'}w_{t'} - \mu_t)(\alpha_{t'}w_{t'} - \mu_t)^T\right)}{trace\left(\Sigma_{t'=1}^{n}(\alpha_{t'}w_{t'} - \mu_t)(\alpha_{t'}w_{t'} - \mu_t)^T\right)}, \quad (7)$$

where α and μ was computed as

$$\alpha_{t'} = \frac{(1/(KL[E_s, E_{t'}] + \varepsilon)^4)}{(1/(KL[E_s, E_i] + \varepsilon)^4)}, \quad (8)$$

$$\mu_t = \frac{1}{n}\sum_{t=1}^{n}\alpha_{t'}w_t. \quad (9)$$

Here, the divergence was calculated by averaging the KL divergences calculated for each class separately. The details of weighted LRTL (wLRTL) approach and the above other algorithms can be reviewed in Azab et al. (2019).

In this study, we discussed the transfer scheme for several subjective conditions of prior knowledge. Inter- and Intra-subject transfer learning approaches were both used to evaluate the methodological effectiveness. Inter-subject transfer calibration trained this classifier with prior experimental trials from other subjects while intra-subject transfer calibration performed this work using its own existing dataset. The target of our research was to find which kind of transfer strategies could be made to improve the online single-trial accuracy of BCI rehabilitation. The right bio-feedback (e.g., robotic arm) was able to raise the subjective confidence and patience to improve the therapeutic effect.

In this experiment, the performances of first sessions were unavailable for intra-subject transfer learning algorithms on

account of no prior dataset. We used the sequential collecting of source data for transfer calibration. That meant that the target session (e.g., Session 5) would be trained by all prior collections (e.g., Session 1, Session 2, Session 3, Session 4). It was consistent with the real condition of model training. Meanwhile, the collection of source data was picked from all other subjects under the inter-subject condition. Specifically, the dataset of each subject which obtained the best performance in all sessions was used for transfer learning. Moreover, 5-fold cross validation was conducted for each approach. All 31 channels of the EEG data were selected for pattern classification. The 45 trials of MA and IS tasks (45 trials per class) were randomly divided into five sets. Four sets were used to train the classifier and the other set was tested to evaluate the performance.

3. RESULTS

3.1. The Experimental Performance of Intra-Subject Based Transfer Learning Approach

In our study, the precision of pattern recognition was considered as the most important index for BCI rehabilitation. **Figure 2** lists the classification results of the above intra-subject transfer learning approaches (LRTL, wLRTL), as well as the baseline approach (LR). The average classification accuracies of all patients were higher than 60% for three different algorithms, except for P6. And it was indicated that the pattern of motor attempt could be distinctive from that of an idle state without motor attempt. However, a paired t-test with Bonferroni correction showed that no discriminatory differences were presented between transfer learning approaches and the baseline approach (LR vs. LRTL: $p = 0.0277$; LR vs. wLRTL: $p = 0.0613$; LRTL vs. wLRTL: $p = 0.6085$). This result suggests that transfer calibration did not significantly improve the BCI performance.

Furthermore, we analyzed the performance of low-precision ($\leq$ 60%) sessions for all patients (**Table 2**). Paired t-test with Bonferroni correction showed that the classification results of transfer learning approaches were significantly greater than that of the baseline approach (LR vs. LRTL: $p = 0.0001$; LR vs. wLRTL: $p = 0.0001$; LRTL vs. wLRTL: $p = 0.0563$). It was meaningfully revealed that transfer calibration could improve the BCI performance induced by poor model training.

3.2. Inter-Subject Based Transfer Learning Approach for MI-Based BCI Rehabilitation

Similarly, **Table 3** lists the classification results of all three algorithms under the inter-subject condition. Paired t-test with Bonferroni correction showed that no discriminative differences were presented between transfer learning approaches and the baseline approach (LR vs. LRTL: $p = 0.0488$; LR vs. wLRTL: $p = 0.1207$; LRTL vs. wLRTL: $p = 0.1744$). This result suggested that the non-significance was consistent with those under the intra-subject condition.

Additionally, low-precision sessions were extracted for further analysis (**Figure 3**). In this case, the statistical analysis indicated

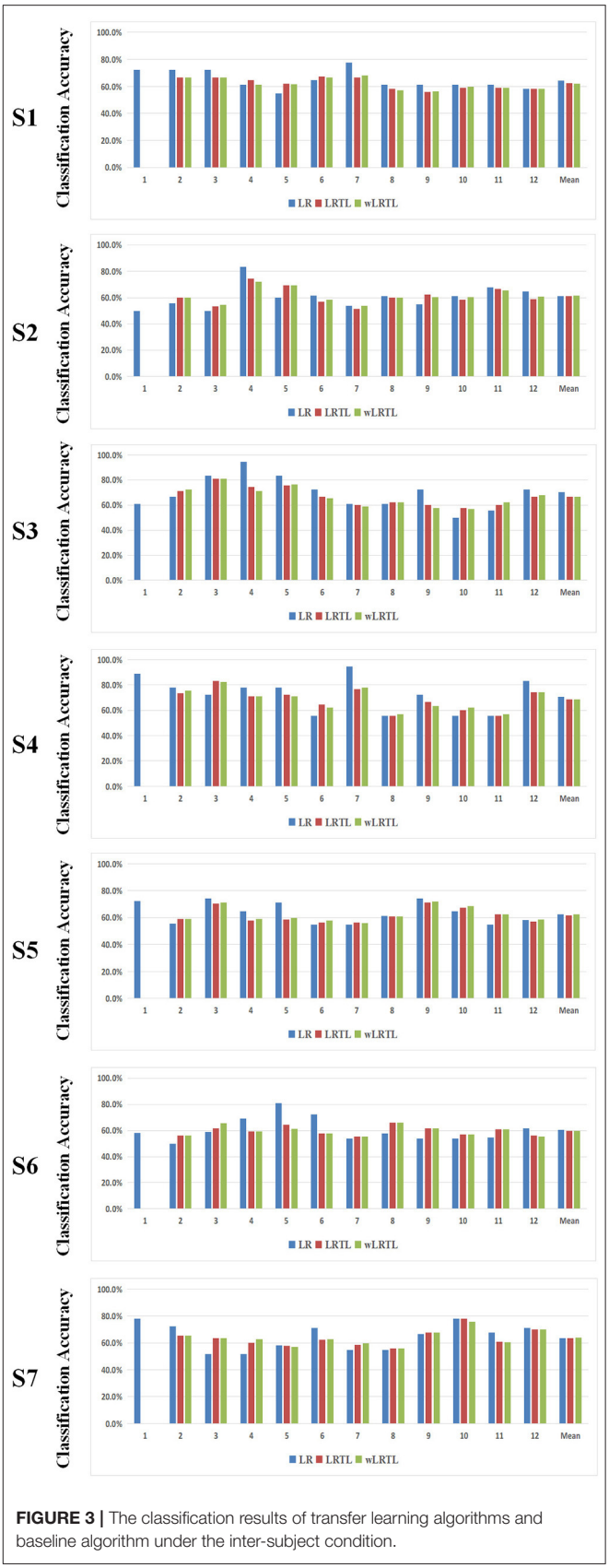


FIGURE 3 | The classification results of transfer learning algorithms and baseline algorithm under the inter-subject condition.

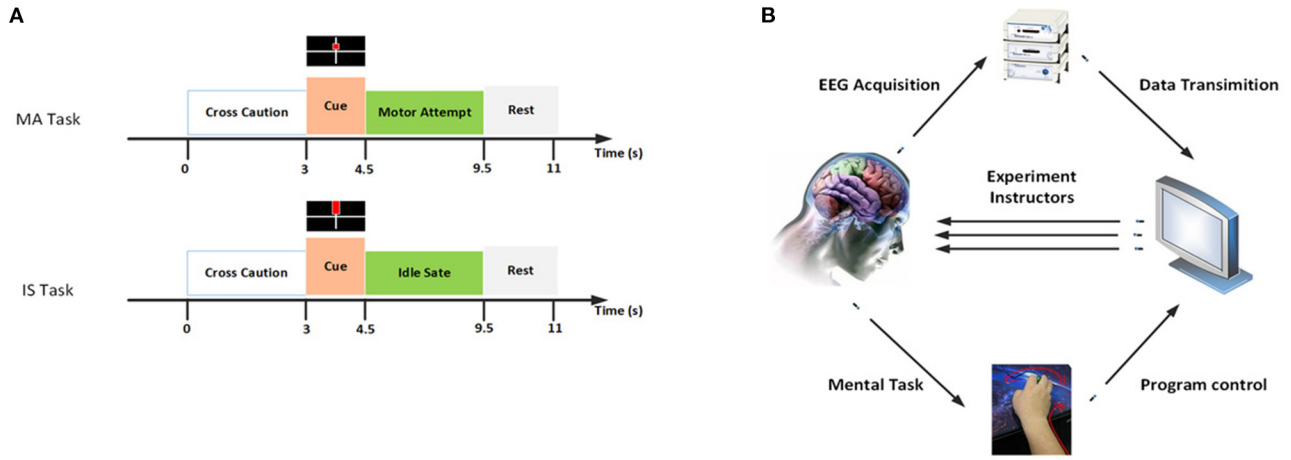


FIGURE 4 | (A) The comparison of low-precision classification results for LRTL algorithm between intra-subject and inter-subject conditions. **(B)** The comparison of low-precision classification results for wLRTL algorithm between intra-subject and inter-subject conditions.

that the accuracies of transfer learning approaches were significantly higher than that of the baseline approach (LR vs. LRTL: $p = 0.0001$; LR vs. wLRTL: $p = 0.0001$; LRTL vs. wLRTL: $p = 0.0210$). It was revealed that transfer calibration improved the BCI performance under the inter-subject condition.

4. DISCUSSION

This study proposed a novel transfer calibration scheme to improve low-precision performance for BCI rehabilitation. This scheme of transfer learning could be used for new subjects without the training dataset, as well as replacing the poor training model–whose accuracy was close to the chance level. Cerebral activities were also observed to clarify the benefit of transfer calibration for feature selection.

4.1. Improvement of Low-Precision Performance for BCI Rehabilitation

As we know, the critical issue of BCI rehabilitation revolves around how to promote the biofeedback effect for active intervention (Ko et al., 2019). It was positive for rehabilitation outcomes, enhancing cortical activity for neural recovery, and increasing confidences of voluntary training (Zhang et al., 2020). Hence, the biofeedback of low-precision trials was negative for patients. Furthermore, most of the sessions used by the baseline classifier achieved the effective performance ($> 60\%$) for each patient, except for S6. It was indicated that very few non-effective results of experimental sessions were unreliable for evaluating this neural treatment. Improved performance of transfer calibration could eliminate the confusion caused by the precision fluctuation.

Moreover, we presented the comparison of transfer learning algorithms between inter-subject (IERS) and intra-subject (IRAS) conditions (**Figure 4**). Statistical analysis was used to compare the classification accuracies among transfer conditions (LRTL: IERS vs. IRAS: $p = 0.3138$; wLRTL: IERS vs. IRAS: $p = 0.3501$;). This result suggested that both of them were reliable for improving low-precision performance of SMR-BCI tasks. However, the advantage of inter-subject transfer calibration was employed for new subjects without training.

The fluctuation of classification for LR resulted from the differences of brain activity between consecutive sessions. It was deduced that changes of cerebral activities were caused by neural self-rehabilitation for patients. As a result, BCI-based intervention was effective for stroke rehabilitation to some extent. Nevertheless, the methodology needs to be further clarified for high-efficiency treatment.

The efficiency of transfer learning has been verified by motor imagery -based BCI tasks for healthy people. However, cerebral impairment of stroke patients would influence the training effect due to the weak neural activities. Therefore, our classification result was lower than those of the above state-of-art BCI systems (Azab et al., 2019, 2020; He and Wu, 2019). Nevertheless, the improvement of low-precision performance was conductive to treatment for the impatient. Compared to the feedback of the random level, the subject was subjectively motivated by positive feedback of right detection. It is meaningful for long-term continuous rehabilitation.

4.2. Transfer Calibration Scheme for SMR-Based BCI Intervention

In our study, an available scheme of transfer calibration was proposed for model selection in the online task of SMR-BCI rehabilitation (**Figure 5**). We summarized several rules as stated below: (1) Instant self-training was necessary for new subjects. The classification model based on current dataset was reliable for BCI tasks. (2) If the patient was frustrated by tedious training, intra- or inter-subject transfer calibration could be used to reduce the calibration time. (3) Furthermore, we could train another model when sufficient trials were finished in the task. If the precision of current model was superior to prior transfer model, the alternative could be automatically performed by our control system. (4) If the model based on the current training dataset

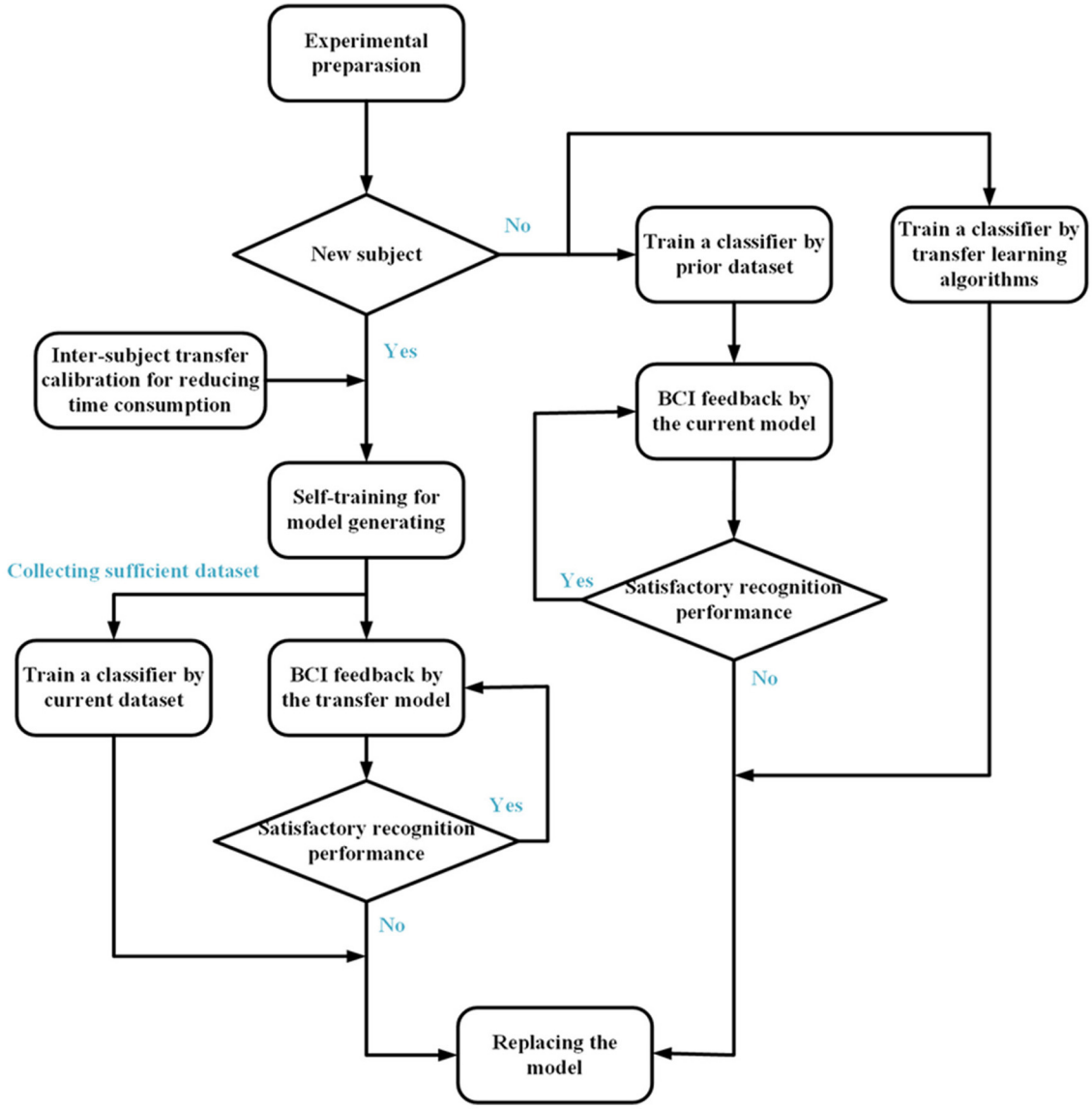


FIGURE 5 | The transfer calibration scheme of classifier model selection for SMR-BCI rehabilitation. For a new subject, intra- or inter-subject transfer calibration could be used for reducing the calibration time. Meanwhile, we could train another model when sufficient trials were finished in the task. If the precision of current using model was superior to that of the spare model, the alternative could be automatically performed by control system.

performed poorly on the experiment, intra- or inter-subject transfer calibration was worth trying in order to replace the under-performing model. (5) For transfer calibration, the volume of EEG data was a crucial factor for model selection between intra-subject and inter-subject conditions. Sufficient training data was an essential precondition for transfer calibration. Specifically, the only option for new subjects without prior experience was inter-subject transfer calibration.

This scheme of transfer calibration was feasible for improving the poor performance of SMR-BCI recognition. And it also reduced the time consumption of model calibration. It was inferred that this scheme was suitable for other transfer learning algorithms.

4.3. Limitation of Current Work

In this study, some issues should be noted and considered in our future work. First, the scheme of transfer calibration needed to be verified for a large amount of stroke patients. It will be addressed in future studies. Second, these patients performed these experiments for 3 months. BCI training over longer time periods should be observed to evaluate performances of the patients in different stages of post-stroke time. Moreover, the number of electrodes was supposed to be reduced by data analysis. It could reduce the time consumption of BCI rehabilitation. Thus, future studies should be conducted to solve these problems to improve the performance of online SMR-BCI rehabilitation.

5. CONCLUSIONS

This paper proposed an effective scheme of transfer calibration for SMR-BCI rehabilitation. The inter- and intra-subject transfer learning approaches could improve the low-precision classification model for BCI feedback. The results imply that this systematical scheme is positive in increasing confidence of voluntary training for stroke patients. It also reduced the time consumption of model calibration.

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 Ethics Committee of Huashan Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LC and HW designed the methodology of data process and performed the data analysis. LC and SC organized the data and wrote the manuscript. LC, ZX, CF, and JJ reviewed and edited the manuscript. All authors read and approved the submitted manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. 61806122), Project of Dai Kerong Academician Workstation (Grant No. MS201758), and Wuxi Health Commission Research Project of China (Grant Nos. ME201944 and T201906).

REFERENCES

Arvaneh, M., Guan, C., Ang, K. K., and Quek, C. (2013). Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain-computer interface. *IEEE Trans. Neural Netw. Learn. Syst.* 24, 610–619. doi: 10.1109/TNNLS.2013.2239310

Azab, A. M., Ahmadi, H., Mihaylova, L., and Arvaneh, M. (2020). Dynamic time warping-based transfer learning for improving common spatial patterns in brain-computer interface. *J. Neural Eng.* 17:016061. doi: 10.1088/1741-2552/ab64a0

Azab, A. M., Mihaylova, L., Ang, K. K., and Arvaneh, M. (2019). Weighted transfer learning for improving motor imagery-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1352–1359. doi: 10.1109/TNSRE.2019.2923315

Azab, A. M., Toth, J., Mihaylova, L. S., and Arvaneh, M. (2018). "A review on transfer learning approaches in brain-computer interface," in *Signal Processing and Machine Learning for Brain-Machine Interfaces*, eds T. Tanaka, and M. Arvaneh (Sheffield: IET), 81–98. doi: 10.1049/PBCE114E_ch5

Birbaumer, N. (2006). Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology* 43, 517–532. doi: 10.1111/j.1469-8986.2006.00456.x

Buch, E., Weber, C., Cohen, L. G., Braun, C., Dimyan, M. A., Ard, T., et al. (2008). Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke. *Stroke* 39, 910–917. doi: 10.1161/STROKEAHA.107.505313

Bundy, D. T., Souders, L., Baranyai, K., Leonard, L., Schalk, G., Coker, R., et al. (2017). Contralesional brain-computer interface control of a powered exoskeleton for motor recovery in chronic stroke survivors. *Stroke* 48, 1908–1915. doi: 10.1161/STROKEAHA.116.016304

Caria, A., da Rocha, J. L. D., Gallitto, G., Birbaumer, N., Sitaram, R., and Murguialday, A. R. (2020). Brain-machine interface induced morpho-functional remodeling of the neural motor system in severe chronic stroke. *Neurotherapeutics* 17, 635–650. doi: 10.1007/s13311-019-00816-2

De Vries, S., Tepper, M., Feenstra, W., Oosterveld, H., Boonstra, A. M., and Otten, B. (2013). Motor imagery ability in stroke patients: the relationship between implicit and explicit motor imagery measures. *Front. Hum. Neurosci.* 7:790. doi: 10.3389/fnhum.2013.00790

Foong, R., Ang, K. K., Quek, C., Guan, C., Phua, K. S., Kuah, C. W. K., et al. (2019). Assessment of the efficacy of EEG-based MI-BCI with visual feedback and EEG correlates of mental fatigue for upper-limb stroke rehabilitation. *IEEE Trans. Biomed. Eng.* 67, 786–795. doi: 10.1109/TBME.2019.2921198

Gao, S., Zhou, M., Wang, Y., Cheng, J., Yachi, H., and Wang, J. (2019). Dendritic neural model with effective learning algorithms for classification, approximation, and prediction. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 601–604. doi: 10.1109/TNNLS.2018.2846646

Grosse-Wentrup, M., Mattia, D., and Oweiss, K. (2011). Using brain-computer interfaces to induce neural plasticity and restore function. *J. Neural Eng.* 8:025004. doi: 10.1088/1741-2560/8/2/025004

He, H., and Wu, D. (2019). Transfer learning for brain-computer interfaces: a Euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* 67, 399–410. doi: 10.1109/TBME.2019.2913914

Hossain, I., Khosravi, A., Hettiarachchi, I., and Nahavandi, S. (2018). Multiclass informative instance transfer learning framework for motor imagery-based brain-computer interface. *Comput. Intell. Neurosci.* 2018, 1–12. doi: 10.1155/2018/6323414

Huang, D., Qian, K., Fei, D.-Y., Jia, W., Chen, X., and Bai, O. (2012). Electroencephalography (EEG)-based brain-computer interface (BCI): a 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 379–388. doi: 10.1109/TNSRE.2012.2190299

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/MCI.2015.2501545

Kang, H., Nam, Y., and Choi, S. (2009). Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Process. Lett.* 16, 683–686. doi: 10.1109/LSP.2009.2022557

Ko, L.-W., Lu, Y.-C., Bustince, H., Chang, Y.-C., Chang, Y., Ferandez, J., et al. (2019). Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface. *IEEE Comput. Intell. Mag.* 14, 96–106. doi: 10.1109/MCI.2018.2881647

Kobler, R. J., and Scherer, R. (2016). "Restricted boltzmann machines in sensory motor rhythm brain-computer interfacing: a study on inter-subject transfer and co-adaptation," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Budapest: IEEE), 469–474. doi: 10.1109/SMC.2016.7844284

Mane, R., Chew, E., Phua, K. S., Ang, K. K., Robinson, N., Vinod, A., et al. (2019). Prognostic and monitory EEG-biomarkers for BCI upper-limb stroke rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1654–1664. doi: 10.1109/TNSRE.2019.2924742

Mane, R., Chouhan, T., and Guan, C. (2020). Bci for stroke rehabilitation: motor and beyond. *J. Neural Eng.* 17:041001. doi: 10.1088/1741-2552/aba162

Meng, J., Zhang, S., Bekyo, A., Olsoe, J., Baxter, B., and He, B. (2016). Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks. *Sci. Rep.* 6:38565. doi: 10.1038/srep38565

Pfurtscheller, G., and Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8

Pichiorri, F., Morone, G., Petti, M., Toppi, J., Pisotta, I., Molinari, M., et al. (2015). Brain-computer interface boosts motor imagery practice during stroke recovery. *Ann. Neurol.* 77, 851–865. doi: 10.1002/ana.24390

Pillette, L., Lotte, F., N'Kaoua, B., Joseph, P.-A., Jeunet, C., and Glize, B. (2020). Why we should systematically assess, control and report somatosensory impairments in BCI-based motor rehabilitation after stroke studies. *Neuroimage Clin.* 28:102417. doi: 10.1016/j.nicl.2020.102417

Ramos-Murguialday, A., Broetz, D., Rea, M., Läer, L., Yilmaz, Ö., Brasil, F. L., et al. (2013). Brain-machine interface in chronic stroke rehabilitation: a controlled study. *Ann. Neurol.* 74, 100–108. doi: 10.1002/ana.23879

Remsik, A. B., Williams, L. Jr., Gjini, K., Dodd, K., Thoma, J., Jacobson, T., et al. (2019). Ipsilesional Mu rhythm desynchronization and changes in motor behavior following post stroke BCI intervention for motor rehabilitation. *Front. Neurosci.* 13:53. doi: 10.3389/fnins.2019.00053

Saha, S., Ahmed, K. I. U., Mostafa, R., Hadjileontiadis, L., and Khandoker, A. (2017). Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 371–382. doi: 10.1109/TNSRE.2017.2778178

Samek, W., Kawanabe, M., and Müller, K.-R. (2013a). Divergence-based framework for common spatial patterns algorithms. *IEEE Rev. Biomed. Eng.* 7, 50–72. doi: 10.1109/RBME.2013.2290621

Samek, W., Meinecke, F. C., and Müller, K.-R. (2013b). Transferring subspaces between subjects in brain-computer interfacing. *IEEE Trans. Biomed. Eng.* 60, 2289–2298. doi: 10.1109/TBME.2013.2253608

Shu, X., Chen, S., Meng, J., Yao, L., Sheng, X., Jia, J., et al. (2018). Tactile stimulation improves sensorimotor rhythm-based BCI performance in stroke patients. *IEEE Trans. Biomed. Eng.* 66, 1987–1995. doi: 10.1109/TBME.2018.2882075

Vidaurre, C., Kawanabe, M., von Bünau, P., Blankertz, B., and Müller, K.-R. (2010). Toward unsupervised adaptation of lda for brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 58, 587–597. doi: 10.1109/TBME.2010.2093133

Wang, Y., Gao, S., and Gao, X. (2006). "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (Shanghai: IEEE), 5392–5395. doi: 10.1109/IEMBS.2005.1615701

Wu, D. (2016). Online and offline domain adaptation for reducing BCI calibration effort. *IEEE Trans. Hum. Mach. Syst.* 47, 550–563. doi: 10.1109/THMS.2016.2608931

Xu, R., Jiang, N., Vuckovic, A., Hasan, M., Mrachacz-Kersting, N., Allan, D., et al. (2014). Movement-related cortical potentials in paraplegic patients: abnormal patterns and considerations for BCI-rehabilitation. *Front. Neuroeng.* 7:35. doi: 10.3389/fneng.2014.00035

Zhang, K., Xu, G., Zheng, X., Li, H., Zhang, S., Yu, Y., et al. (2020). Application of transfer learning in EEG decoding based on brain-computer interfaces: a review. *Sensors* 20:6321. doi: 10.3390/s20216321

Zhao, X., Zhao, J., Cai, W., and Wu, S. (2019). Transferring common spatial filters with semi-supervised learning for zero-training motor imagery brain-computer interface. *IEEE Access* 7, 58120–58130. doi: 10.1109/ACCESS.2019.2913154

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 04 February 2021
doi: 10.3389/fnins.2021.614182

GaborNet Visual Encoding: A Lightweight Region-Based Visual Encoding Model With Good Expressiveness and Biological Interpretability

*Yibo Cui, Kai Qiao, Chi Zhang, Linyuan Wang, Bin Yan and Li Tong**

Henan Key Laboratory of Imaging and Intelligent Processing, PLA Strategic Support Force Information Engineering University, Zhengzhou, China

Computational visual encoding models play a key role in understanding the stimulus–response characteristics of neuronal populations in the brain visual cortex. However, building such models typically faces challenges in the effective construction of non-linear feature spaces to fit the neuronal responses. In this work, we propose the GaborNet visual encoding (GaborNet-VE) model, a novel end-to-end encoding model for the visual ventral stream. This model comprises a Gabor convolutional layer, two regular convolutional layers, and a fully connected layer. The key design principle for the GaborNet-VE model is to replace regular convolutional kernels in the first convolutional layer with Gabor kernels with learnable parameters. One GaborNet-VE model efficiently and simultaneously encodes all voxels in one region of interest of functional magnetic resonance imaging data. The experimental results show that the proposed model achieves state-of-the-art prediction performance for the primary visual cortex. Moreover, the visualizations demonstrate the regularity of the region of interest fitting to the visual features and the estimated receptive fields. These results suggest that the lightweight region-based GaborNet-VE model based on combining handcrafted and deep learning features exhibits good expressiveness and biological interpretability.

Keywords: GaborNet-VE, visual encoding model, fMRI, expressiveness, biological interpretability

OPEN ACCESS

Edited by:
Kaijian Xia, Changshu No.1 People's Hospital, China

Reviewed by:
Yalin Wang, Arizona State University, United States
Yuanqing Li, South China University of Technology, China

***Correspondence:**
Li Tong
tttocean_tl@hotmail.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *05 October 2020*
Accepted: *04 January 2021*
Published: *04 February 2021*

Citation:
Cui Y, Qiao K, Zhang C, Wang L, Yan B and Tong L (2021) GaborNet Visual Encoding: A Lightweight Region-Based Visual Encoding Model With Good Expressiveness and Biological Interpretability. Front. Neurosci. 15:614182. doi: 10.3389/fnins.2021.614182

INTRODUCTION

Human and primate visual systems are exceedingly adept at achieving complicated vision tasks based on rudimentary visual perception. A key research goal in the study of such visual systems is to comprehensively understand the neuronal basis of vision, especially the stimulus–response characteristics of neuronal populations (Ukita et al., 2019). This goal can be achieved through using functional magnetic resonance imaging (fMRI), which can measure brain activity from human subjects passively viewing natural images (Naselaris et al., 2011; Kay, 2018). In particular, based on the blood-oxygen-level-dependent (BOLD) signals, visual encoding models can be built to predict neural responses to arbitrary stimuli and hence develop a better understanding of visual information processing in the human brain (Cadena et al., 2019). However, a key challenge in this modeling and prediction process is the difficulty of constructing an effective non-linear feature space to analyze the non-linearity of neuronal responses.

In early visual encoding models, feature spaces were mainly constructed based on non-linear handcrafted features inspired by some visual representation mechanisms. As well, previous studies showed that neurons in the early stages of cortical visual processing have smaller receptive fields (RFs) and higher sensitivity to low-level features such as texture (Hubel and Wiesel, 1962), whereas neurons further down the ventral stream have larger RFs, higher representation invariance, and stronger response to complicated shapes (Hubel and Wiesel, 1968; Hung et al., 2005). In addition, neural mechanisms in the primary visual regions could be reliably modeled by Gabor wavelets with variations in location, orientation, and spatial frequency (Adelson and Bergen, 1985; Jones and Palmer, 1987; Carandini et al., 2005). Hence, Kay et al. (2008) proposed the Gabor wavelet pyramid (GWP) visual encoding model that consists of an over-complete basis of phase-invariant Gabor wavelets with different positions, orientations, and spatial frequencies. This model processes visual stimuli to generate a non-linear feature space with good expressiveness and interpretability. Thus, the GWP model became a classical encoding model for low-level visual cues. The key aspects and prediction outcomes of the GWP model were further improved in many subsequent studies (Naselaris et al., 2009; Nishimoto et al., 2011; Vu et al., 2011; Li et al., 2018; St-Yves and Naselaris, 2018). Nevertheless, the GWP model is not suitable for higher-level visual processing. Naselaris et al. (2009) made a voxel-to-voxel comparison of natural scene category labels between the GWP model and a semantic model. The results demonstrated that although the predictions of the two models are competitive, each model excels the other. Specifically, the GWP model predictions are more accurate for voxels in the primary visual cortex, whereas the semantic model predictions show higher accuracy for voxels in the secondary visual cortex. Because high-level features are hard to design manually, the feature space is usually composed of semantic labels obtained through manual image annotation. Consequently, the voxel activity in several secondary cortex areas could be accurately predicted by semantic features (Mitchell et al., 2008; Naselaris et al., 2009, 2012; Huth et al., 2012; Stansbury et al., 2013). However, the performance of the semantic models for visual encoding is typically degraded by the inevitable subjective selection and judgment bias in the manual semantic labeling process. In summary, encoding models based on handcrafted features have strong biological interpretability but limited prediction performance and poor universality because such models are usually designed for few certain visual regions.

Recently, deep neural networks (DNNs) have made breakthroughs in a variety of domains (such as computer vision) and hence caught the attention of researchers in computational neuroscience (Kietzmann et al., 2017; Cichy and Kaiser, 2019; Richards et al., 2019; Serre, 2019). Indeed, numerous methods in neuroscience have used convolutional neural networks (CNNs) to model the human visual system and thus achieved unprecedented improvements in creating visual encoding models. Generally, these CNN-based visual encoding modeling approaches can be categorized into either task- or data-driven approaches.

On the one hand, the task-driven approaches essentially model neural responses by generating non-linear features based on an intermediate layer of a CNN pretrained on a higher vision task (Yamins and DiCarlo, 2016). This transfer learning approach has boosted the performance in many computer vision and machine learning tasks where labeled data are limited (Oquab et al., 2014). Motivated by these performance improvements, Agrawal et al. (2014) used CNN-based features to model brain activity and simultaneously predict such activity with high accuracy in the low-level, intermediate, and high-level stages of the visual pathway. Güçlü et al. used two different CNNs for layer-wise analysis of voxel scores in the ventral stream (Güçlü and van Gerven, 2015) and dorsal stream (Güçlü and van Gerven, 2017). The experimental results of these two methods reveal a gradient in the representation complexity. More recently, feature representations have been improved through novel DNN architectures, such as ResNet (Wen et al., 2018), recurrent neural networks (Shi et al., 2018), and the variational autoencoder (Han et al., 2017). However, these feature representations were not directly learned for the visual encoding task, a shortcoming that resulted in a data–task mismatch. Hence, the associated fixed features are not suitable for encoding visual cortical responses. Moreover, from a machine learning perspective, a task-driven approach has two independent linear and non-linear mapping components, and the corresponding optimization scheme might fail to reach the global optimum and instead get stuck into a local optimum.

On the other hand, the data-driven approaches directly learn all parameters of a designed encoding model based on experimental stimuli to obtain visual cortical responses. This end-to-end scheme led to successful DNN applications in many domains and thus replaced the traditional two-step machine-learning scheme. For example, Qiao et al. (2019) designed an end-to-end CNN regression (ETECR) model for visual encoding based on fMRI data. This model combined linear and non-linear mapping components into the CNN architecture, which was trained by experimental stimuli and corresponding fMRI signals. Hence, this model could learn optimal feature representations and linear regression weights for visual cortical responses and achieve major improvements in prediction performance. However, the ETECR model has limited biological interpretability because of the black box nature of its CNN underlying architecture.

Collectively, visual encoding models based on handcrafted features have good biological interpretability, but their prediction performance is limited. Alternately, visual encoding models based on CNN features exhibit excellent expressiveness but lack sufficient biological interpretability. Therefore, hybrid visual encoding models are sought to combine the biological interpretability of handcrafted features and the expressiveness of deep features. In this paper, we introduce the end-to-end GaborNet visual encoding (GaborNet-VE) model for the brain visual ventral stream. This model is trained on non-linear deep feature representations based on Gabor features (Liu and Wechsler, 2001, 2002; Kyrki et al., 2004) (to be introduced in detail in section "Gabor Filters in the First GaborNet Visual Encoding Layer"). Firstly, we designed a

lightweight GaborNet-VE regression model composed of a Gabor convolutional layer, two regular convolution layers, and a fully connected layer. The Gabor convolutional layer embeds a set of learnable parametric Gabor filters. The GaborNet-VE model has an efficient region-based encoding scheme where all voxels in one region of interest (ROI) are jointly encoded. Selective optimization of features and voxels was used in our model to weigh them more effectively. In comparison with three reference models (namely, the GWP, CNN-linear, and ETECR models), our model achieved state-of-the-art prediction performance for the primary visual cortex and comparable prediction performance for the intermediate and higher cortex areas. Hence, the GaborNet-VE model has good expressiveness. Then, the guided backpropagation (GBP) algorithm was used to visualize the effective pixels and the Gabor kernels for each voxel based on the response of the top 100 best-predicted voxels in each ROI. Our visualizations showed that the proposed visual encoding model could measure the RFs of each voxel. In fact, voxels in early visual areas had smaller RFs, whereas voxels further down the ventral pathway had larger RFs. Moreover, our model could learn the preferred Gabor kernels of each voxel. The properties of these kernels demonstrated the properties of fitting to visual features of ROI. Overall, the voxels in V1 and V2 prefer Gabor kernels with high spatial frequencies, whereas the voxels in V4 and LO prefer Gabor kernels with low spatial frequencies, although special voxels with opposite properties exist in all areas. These results suggest that the lightweight GaborNet-VE model based on combining handcrafted and deep ROI features has both good expressiveness and biological interpretability. Clearly, our work connects deep learning with neuroscience and promotes the development of artificial intelligence and the understanding of human intelligence.

MATERIALS AND METHODS

Experimental Data

To evaluate the proposed methods, a publicly available dataset (introduced in Kay et al., 2008 and Naselaris et al., 2009) was analyzed. All experimental details were presented in the studies mentioned earlier (Kay et al., 2008; Naselaris et al., 2009). Hence, we will only briefly summarize the data collection process here.

The dataset contains training and validation data of BOLD fMRI responses preprocessed after being collected from two male subjects (S1 and S2), while they were viewing natural images. The training image library included 1,750 grayscale images, each of which was presented twice. The validation image library included 120 different grayscale images, each of which was presented 13 times. Photographs were presented in successive 4-s trials; in each trial, a photograph was presented for 1 s, and the gray background was presented for 3 s. Each 1-s presentation consisted of a photograph being flashed ON–OFF–ON–OFF–ON where ON corresponds to the presentation of the photograph for 200 ms and OFF corresponds to the presentation of the gray background for 200 ms. The data collection was performed using a 4-T Varian INOVA magnetic resonance (Varian, Inc., Palo Alto, CA, United States) scanner. Eighteen coronal slices were obtained from the occipital cortex (slice thickness = 2.25 mm, slice gap = 0.25 mm, field-of-view = 128 mm × 128 mm). The BOLD signals were collected using a T2*-weighted, slice-interleaved, single-shot, gradient-echo pulse sequence of echo-planar imaging (spatial resolution = 2 mm × 2 mm × 2.5 mm, flip angle = 20°, TE = 28 ms, TR = 1 s, matrix size 64 × 64).

Framework of the GaborNet Visual Encoding Model

The GaborNet-VE model adopted a region-based encoding scheme where all voxels of one ROI in the visual cortex are jointly encoded (Zhang et al., 2019). Therefore, the GaborNet-VE model was trained by fMRI data collected from all ROIs. Model training and testing were implemented under the deep learning framework, PyTorch (0.4.0). The proposed visual encoding model consists of an input layer, a Gabor convolutional layer, several regular convolutional layers, several fully connected layers, and an output layer. The activation function of a convolutional layer is defined as the rectified linear unit (ReLU) (Nair and Hinton, 2010) transformation of a two-dimensional convolution of the activation function of the previous layer. The activation function of a fully connected layer is defined as the non-linear ReLU transformation of the weighted sum of the activation functions of the previous layer. In addition, the Gabor convolutional layer can be of a one-way real type only or a two-way combination of real and imaginary types. The number of Gabor convolutional filters in that layer is 64 or 128, and the size of each filter is (7, 7), (9, 9), (11, 11), or (13, 13). The number of convolutional filters in a regular convolutional layer is 64, and the size of each filter is (3, 3). In each convolutional layer, the stride size is (2, 2), and valid padding is used. The mini-batch size, the optimizer type [stochastic gradient descent or adaptive moment estimation (Adam) (Kingma and Ba, 2014)], the learning rate decay coefficient, and the number and types of hidden layers (Gabor convolutional, regular convolutional, or fully connected layers) were optimized with a fivefold cross-validation scheme for the fMRI data of V1 and V2. The optimized hyperparameters of GaborNet-VE are as follows: the hidden layer structure has one two-way Gabor convolutional layer and two regular convolutional layers, followed by one fully connected layer (**Figure 1**); the number of Gabor convolutional filters is 128 (equally divided among the real and imaginary types); the size of each Gabor filter is (9, 9); the mini-batch size is 128; the Adam optimizer is used; the learning rate decay coefficient is 0.001. All other hyperparameters were kept fixed. Moreover, GaborNet-VE used the ETECR model learning strategy with selective optimization of features and voxels. Details of this learning strategy are shown in section "Optimization Strategy."

Gabor Filters in the First GaborNet Visual Encoding Layer

The Gabor filters, introduced by Dennis Gabor, is a family of bandpass filters, which accept or reject inputs within a range of spatial frequencies (Gabor, 1946). These filters have been used as an efficient tool in diverse pattern analysis

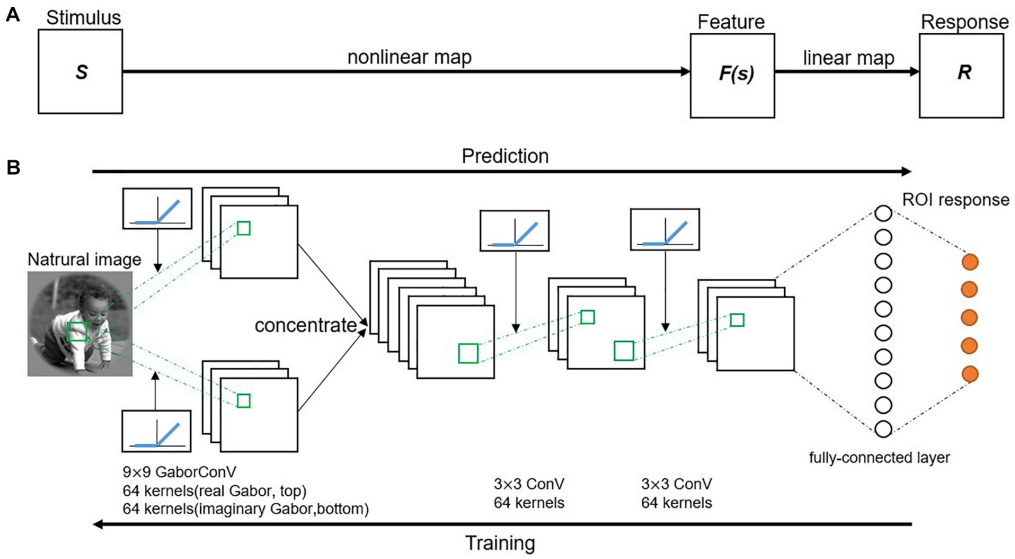


FIGURE 1 | Proposed visual encoding model. **(A)** Model framework. A visual stimulus is transformed into a voxel response in two stages. First, visual stimulus (S) is transformed into a non-linear feature space [F(s)] by a non-linear mapping. Then, feature space is transformed into a voxel response (R) by a linear mapping. **(B)** A schematic diagram of the GaborNet-VE model. Response to a natural image is predicted by an end-to-end model consisting of a Gabor convolutional layer, two successive regular convolutional layers, and a fully connected layer. Gabor convolutional layer has 128 Gabor kernels with 64 kernels for each of the real and imaginary types. Size of each Gabor kernel is 9 × 9. Also, each regular convolutional layer has 64 kernels of a size of 3 × 3. Each convolutional layer is followed by a rectified linear unit (ReLU) transformation. Fully connected layer calculates weighted sum of inputs from previous layer followed by a ReLU transformation.

applications (Huang et al., 2004) for extracting different types of textures, edges, and spatially localized spectral features. Recent deep network visualization results demonstrated that Gabor-like kernels were mostly used in the first convolutional layers of CNNs trained on large-scale natural image datasets (Krizhevsky et al., 2017). The similarity between Gabor and convolutional kernels and the inherent error resiliency of deep networks represent the basis for incorporating Gabor kernels into the proposed network. Hence, a Gabor convolutional layer replaces the regular convolutional layer as the first layer in our GaborNet-VE model.

A Gabor filter is a Gaussian filter modulated by a complex sinusoidal wave (Alekseev and Bobe, 2019). This filter is monotonous and differentiable and can be defined mathematically as follows:

$$g\left(x, y; \omega, \sigma, \varphi, \theta\right) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \times exp(j\left(\omega x' + \varphi\right)), \tag{1}$$

$$g_{rea}\left(x, y; \omega, \sigma, \varphi, \theta\right) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \times cos\left(\omega x' + \varphi\right), \tag{2}$$

$$g_{img}\left(x, y; \omega, \sigma, \varphi, \theta\right) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \times sin\left(\omega x' + \varphi\right), \tag{3}$$

Where

$$x' = x\,cos\theta + y\,sin\theta, \tag{4}$$

$$y' = -x\,sin\theta + y\,cos\theta. \tag{5}$$

The complex form of the Gabor filter (Eq. 1) can be decomposed into a real part (Eq. 2) and an imaginary part (Eq. 3). The center frequency of the Gabor wavelet is controlled by Eq. 4 and Eq. 5. The Gabor filter can be modified through the tuning of many parameters, which plays various roles in extracting image features. The standard deviation parameter, σ, within the Gaussian function controls the function spread. The orientation parameter within the sinusoidal wave is denoted by θ, and is used by the filter to extract features at different angles. The sinusoidal spatial frequency parameter, ω, controls the wavelength and the bar width for a Gabor wavelet. Indeed, a Gabor wavelet with a wide bar has a low ω, whereas the one with a narrow bar has a large ω. The phase shift parameter, φ, controls the sinusoidal phase offset. These parameters are initialized based on a previous study (Meshgini et al., 2012). In particular, the ω and θ parameters of the Gabor kernels are set as follows:

$$\omega_m = \frac{\pi}{2}\sqrt{2}^{m},\ m = 0, 1, \ldots, 4, \tag{6}$$

$$\theta_n = \frac{\pi}{8}(n-1),\ n = 1, 2, \ldots, 8. \tag{7}$$

The σ and φ parameters are initialized randomly from the uniform distributions U (0, 5) and U (0, π), respectively.

Optimization Strategy

Following the ETECR model (Qiao et al., 2019), we adopt a selective optimization strategy of features and voxels in our model. To increase voxel attention to well-related features, the fully connected layer used self-adapting regression weights, a policy that is different from that of general linear regression

with regularization. This dynamical learning scheme of feature weights was implemented by squaring the original weights. In Eq. 8, one stimulus image, s_i, is transformed into the corresponding feature space (f_i) by three convolutional layers, $Conv(\,)$.

$$f_i = Conv(s_i), \tag{8}$$

In Eq. 9, a regular weight w_{fc} is replaced by its square w_{fc}^2 in the fully connected layer, which maps the features to the predicted responses, $\widehat{r_i}$.

$$\widehat{r_i} = w_{fc}^2 f_i + b, \tag{9}$$

In this way, the learning rate (μw_{fc}) of weights can be dynamically adjusted according to the current status, where μ is the original learning rate,

$$\Delta w_{fc} = -\mu w_{fc} \Delta \widehat{r_i}, \tag{10}$$

For selective voxel optimization, noise regularization, and a weighted correlation loss function were used. To reduce the influence of ineffective voxels, Gaussian noise n_g with zero mean and unit variance is added to each of the predicted responses $\widehat{r_i}$,

$$\widehat{r_i} = \widehat{r_i} + n_g, \tag{11}$$

The Pearson correlation ($\rho_m \in [-1, 1]$) of the predicted responses (r_m) and the actual responses ($\widehat{r_m}$) of the m_{th} voxel is calculated as

$$\rho_m = cor(r_m, \widehat{r_m}) = \frac{Cov(r_m, \widehat{r_m})}{\sqrt{Var(r_m) \cdot Var(\widehat{r_m})}}, \tag{12}$$

To focus the attention of the proposed model on more effective voxels, we introduced the dynamical weights, $\eta_m = \rho_m^2$, in the loss function

$$Loss = -\frac{\sum_m \eta_m \rho_m}{n} + \gamma \left| \frac{\sum_m \widehat{r_m}}{n} \right|, \tag{13}$$

instead of computing the average ρ of all voxels in one ROI. The final form of the loss function is given by

$$Loss = -\frac{\sum_m \rho_m^3}{n} + \gamma \left| \frac{\sum_m \widehat{r_m}}{n} \right|, \tag{14}$$

where γ is used to control the relative contributions of the fidelity and regularization terms.

Guided Backpropagation

With the emergence of deep network architectures, several algorithms have been proposed to visualize and interpret network outcomes. Firstly, a deconvolutional network (DeconvNet) approach was used to visualize the most discriminative image details for given neurons in network layers (Zeiler and Fergus, 2013). However, the DeconvNet visualization results are not generally clear and recognizable, especially with the large computational complexity of high-level features. An alternative visualization approach based on backpropagation (Rumelhart et al., 1986) involves computing the backward gradient of the activation function of a single unit. The main difference between the backpropagation and deconvolution approaches is the method of calculating the backward signal passing through the ReLU non-linearity. Although the backpropagation approach handles only backward values with gradients greater than 0,

$$G_i^l = \left(f_i^l > 0\right) \times G_i^{l+1}, \tag{15}$$

the DeconvNet approach deals only with backward values greater than 0,

$$G_i^l = \left(G_i^{l+1} > 0\right) \times G_i^{l+1}, \tag{16}$$

where f_i^l is a bottom feature in the layer l from the forward direction, and G_i^l is a top gradient in the layer l from the backward direction. The GBP visualization algorithm (Springenberg et al., 2014) combines backpropagation and deconvolution approaches and masks out both negative values corresponding to the top gradient and bottom feature,

$$G_i^l = \left(f_i^l > 0\right) \times \left(G_i^{l+1} > 0\right) \times G_i^{l+1}. \tag{17}$$

The GBP algorithm enhances the performance of the ordinary backpropagation algorithm through limiting negative gradients. Such negative gradients are indeed undesirable, as they are typically associated with image regions that weaken feature visualization.

Reference Visual Encoding Models

To further assess the performance of our proposed model, we compared it with the following three state-of-the-art visual encoding models:

Gabor Wavelet Pyramid Model (Kay et al., 2008)

The GWP model is a classical voxel-based visual encoding model proposed by Kay et al. (2008). This model has demonstrated superior performance in visual retrieval tasks based on fMRI. The model firstly processes stimulus images using a family of handcrafted quadrature-phase Gabor wavelets with different spatial frequencies, orientations, and locations. Then, the prediction responses are mapped linearly by the square roots of the pooled energies of the Gabor features. In our experiments, the same number of Gabor wavelets as that in Kay et al. (2008) was used on 128 × 128-pixel images, and a linear mapping was obtained by regularized orthogonal matching pursuit (Zhang et al., 2019).

Convolutional Neural Network-Linear Model

The CNN-linear model has several advantages, which have resulted in remarkable visual encoding performance. The model used herein follows the description in Güçlü and van Gerven (2015) and Eickenberg et al. (2016). Firstly, the feature space is composed of non-linear features obtained from each of the eight layers of the pretrained AlexNet model. Then, a linear mapping from the feature space to the brain activity space is fitted using the regularized orthogonal matching pursuit algorithm. Hence, eight voxel-based encoding models were constructed from eight feature spaces for each ROI voxel. Lastly, the model with the best prediction performance was selected as the CNN-linear encoding model.

End-to-End CNN Regression Model (Qiao et al., 2019)

The two encoding models mentioned earlier are linearized voxel-based models with a two-step procedure. To overcome the limitations of these models, the ETECR model was proposed. The ETECR model has four convolutional layers and one fully connected layer and is trained by an end-to-end regression scheme. The self-adapting weight learning and weighted correlation loss are used in the model construction, leading to streamlined optimization of the effective voxels. Moreover, the ETECR model encodes jointly and efficiently all voxels of one ROI. Therefore, the ETECR model simultaneously offers a lightweight architecture, effectiveness, and efficiency.

Quantification of Model Performance

To quantify the prediction performance of each encoding model, the voxel-wise prediction accuracy is defined as the Pearson correlation (see Eq. 12) between the real and predicted responses on the validation dataset. For each model type, we carried out a replacement test to test whether the prediction accuracy of each voxel is significantly bigger than the value set by the null hypothesis. For the 120 images in the validation dataset, we randomly disturbed the correspondence between the real and predicted responses and recalculated the prediction accuracy of each voxel. After repeating this procedure 1,000 times, a null-hypothesis distribution was generated. Then, the response for one voxel can be predicted accurately when the prediction accuracy is higher than 0.27, which is significantly above the null hypothesis distribution ($p < 0.001$). Then, the replacement test was used to assess the significance of a model superiority compared with the other models. The voxels, whose responses were accurately predicted by both models, were selected for random shuffling with a 0.5 probability. According to the null-hypothesis distribution obtained by repeating the procedure mentioned earlier 1,000 times, one model significantly outperforms the other when the former can accurately predict more than 53% of the voxels ($p < 0.05$).

To compare the visual encoding models for a group of voxels, we considered two approaches as follows. In the first approach, a voxel-to-voxel comparison is made for the same voxel among the two compared models. As shown in **Figure 2**, scatter plots were used for visualizing and clarifying the results. In these plots, each dot represents one individual voxel. Therefore, the best encoding model for a single voxel can be readily identified. Then, the voxels in each interval were counted to construct an estimate of the voxel density in bar plots (see **Figure 2**). The second approach compares the overall model performance outcomes after sorting all voxels. As shown in **Figure 3**, line plots were constructed to represent voxels in descending order of the prediction accuracy.

RESULTS

We trained, evaluated, and then tested the GaborNet-VE model using a subset of a classical dataset, which contained the functional BOLD signals generated in response to viewing grayscale natural images. The signals were generated from 1,294, 2,083, 1,790, 1,535, and 928 voxels in the V1, V2, V3, V4, and LO areas in the ventral visual stream of S1, respectively. The GaborNet-VE model followed a region-based encoding scheme, where all voxels of one ROI are jointly encoded. Therefore, five GaborNet-VE models with different parameters were fit by the stimuli and fMRI data of the five areas.

Gabor Kernel Visualization and Receptive Field Estimation

The Gabor kernels in the first convolutional layer of each of the five constructed GaborNet-VE models are visualized in **Figure 4A**. It is quite challenging to analyze the similarities and differences of the five visual areas from the Gabor kernel maps and to find out which Gabor kernel plays an essential role for specific voxels. Therefore, we implemented the GBP algorithm to back-propagate from the top 100 voxels with the best predictions in each visual area to obtain the RF on the stimulus image and eight preferred Gabor kernels for each voxel. For convenience and space limitations, only five representative voxels were selected from the 100 voxels of each visual area to visually summarize the overall results (see **Figure 4B**). In **Figure 4B**, "Vol." is defined as a form of the serial number of a voxel. For V1, the RFs of most voxels are concentrated in a small region, and these voxels preferred Gabor kernels with smaller scales and higher spatial frequencies (e.g., Vol. 5, Vol. 178, Vol. 689, and Vol. 826). However, a small number of voxels preferred larger Gabor kernels with lower spatial frequencies (such as Vol. 870, which has a significantly large RF). For V2, some voxels exhibited RFs similar to those of V1 (e.g., Vol. 363 and Vol. 383), whereas most V2 voxels show larger concentrated RFs than those of the V1 voxels. The preferred Gabor kernels of the V2 voxels (such as Vol. 300 and Vol. 321) have larger scales and lower spatial frequencies than those of the V1 voxels. However, a few voxels prefer Gabor kernels with larger scales and lower spatial frequencies (such as Vol. 490, whose RF is relatively large and more scattered). For V3, most voxels have larger RFs than those of the V2 voxels, but some of these V3 voxels (such as Vol. 65 and Vol. 81) are concentrated, and some others (such as Vol. 1445) are dispersed. There are a few V3 voxels, such as Vol. 284 and Vol. 410, with similar RFs to those of the voxels in V1 or V2. In addition to the irregular pattern of the RFs, the preferred Gabor kernels of the voxels in V3 are also more complex, with various scales and spatial frequencies. For V4, many voxels (such as Vol. 111 and Vol. 132) have larger concentrated RFs than those of the voxels in V3. Also, there are a few V4 voxels, such as Vol. 1397, with small RFs such as those of the voxels in V1 or V2. There are voxels, such as Vol. 257, with RFs that cover most of the image and that begin to appear for the first time in V4. For LO, there are some voxels (e.g., Vol. 76 and Vol. 145) with large RFs such as those of the voxels in V4. Other LO voxels (e.g., Vol. 374 and Vol. 427) have small RFs such as those of the voxels in V1 or V2, whereas a few voxels (e.g., Vol. 253) have RFs that cover almost the whole image. The preferred Gabor kernels of the voxels in V4 and LO are similar, and these kernels mostly have large scales and low spatial frequencies.

In summary, most of the voxels in V1 and V2 have small concentrated RFs, whereas most of the voxels in V4 and LO have

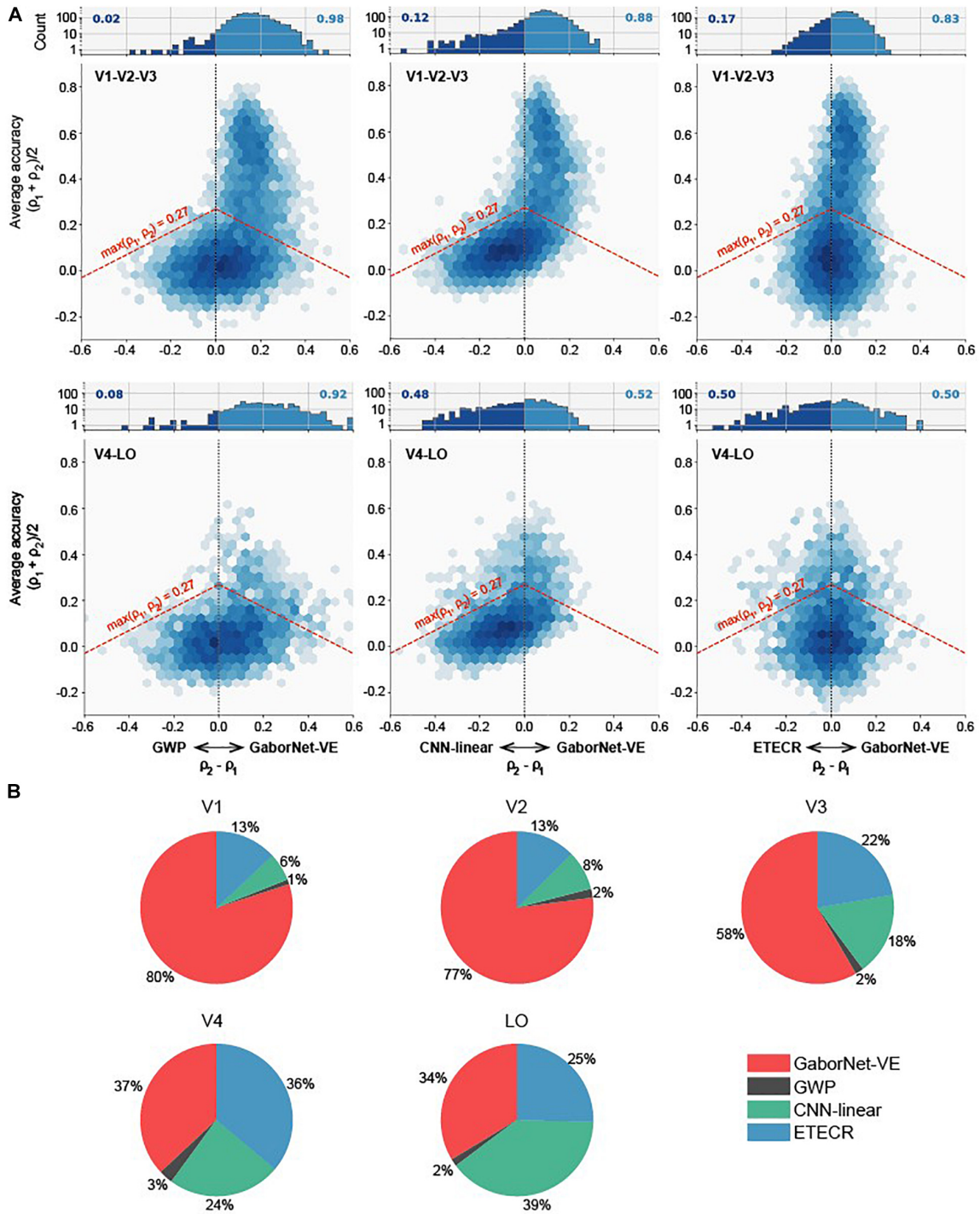


FIGURE 2 | Voxel-to-voxel comparisons of GaborNet-VE, GWP, CNN-linear, and ETECR models. **(A)** Pairwise comparisons of GaborNet-VE model the GWP, CNN-linear, and ETECR models. Each of six performance plots shows a comparison of prediction accuracies for two models. Position along vertical axis indicates average prediction accuracy for models under comparison; shifts to right or left along the horizontal axis indicate a relative improvement in prediction accuracy for one of compared models. Color of each hexagonal bin indicates log-scaled number of voxels in a local region of plot. Histogram at top of each plot represents distribution of relative improvements for all voxels whose prediction accuracy is above 0.27 ($p < 0.001$, randomization test) for at least one of two models. This distribution corresponds graphically to all voxels above red dashed line. Number on each side represents fraction of voxel predictions that are improved under corresponding model. In all plots, a shift of the data toward right indicates an advantage for GaborNet-VE model, whereas a shift toward left indicates an advantage for reference models. Upper plots display data for voxels in primary visual cortex (V1, V2, and V3); lower plots display data for voxels in intermediate and higher visual areas (V4 and LO). **(B)** Joint comparisons of GaborNet-VE model and all of reference models. Each plot contains percentages of voxels in one ROI best predicted by each encoding model. Red, gray, green, and blue curves represent GaborNet-VE, GWP, CNN-linear, and ETECR models, respectively.

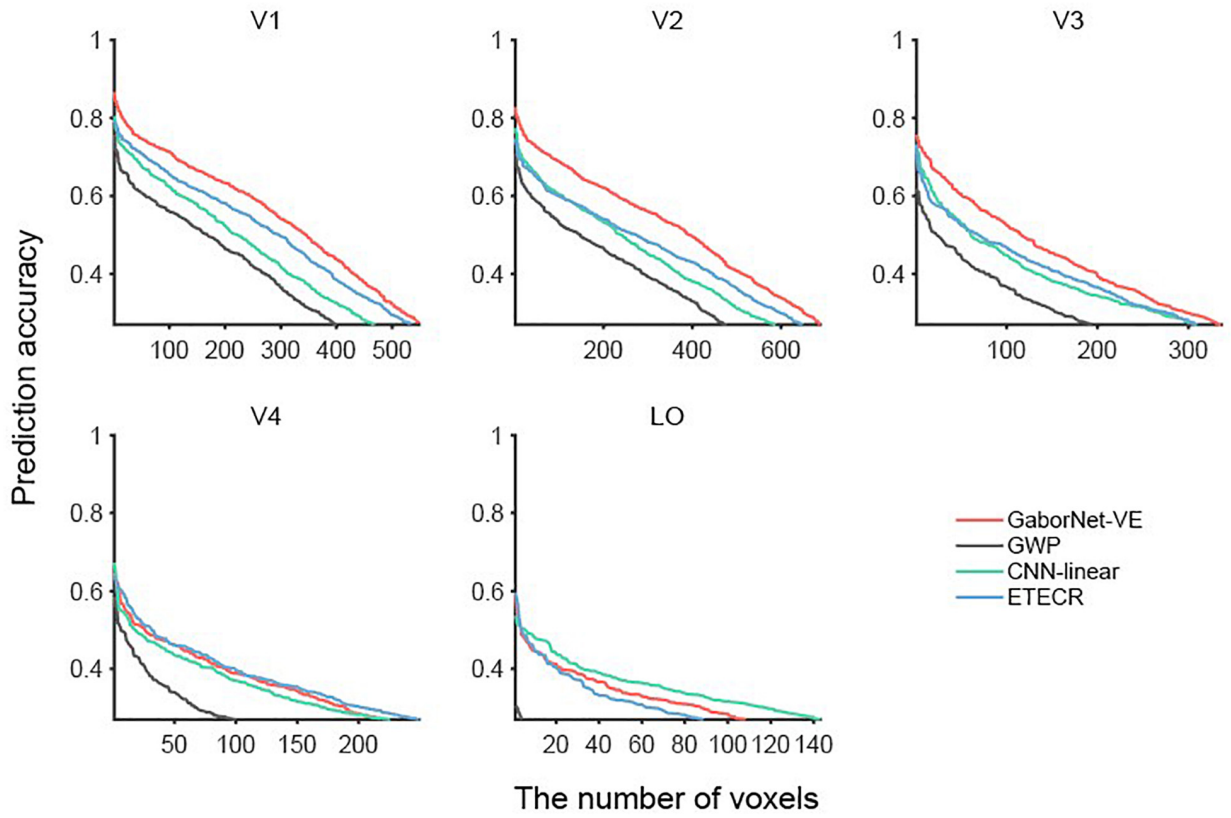


FIGURE 3 | Comparisons of GaborNet-VE, GWP, CNN-linear, and ETECR models by sorting voxels in descending order of prediction accuracy. Only accurately predicted voxels are plotted. Red, gray, green, and blue curves represent GaborNet-VE, GWP, CNN-linear, and ETECR models, respectively.

large concentrated RFs. The preferred Gabor kernels of most of the voxels in V1 and V2 have small scales and high spatial frequencies, whereas the kernels of most of the voxels in V4 and LO have large scales and low spatial frequencies. However, a few voxels oppose these conclusions for V1, V2, V4, and LO. Moreover, the voxels in V3 are capricious, regardless of the size of the RF and the properties of the preferred Gabor kernels. To show the overall results more clearly, we have computed statistics for the results of the top 100 voxels of each area. The average ratio of the RF size to the entire image size is plotted in **Figure 4C**. Obviously, the RF size becomes larger as we go from V1 to V4 but decreases as we go from V4 to LO. The means of the four parameters ($\omega, \sigma, \varphi, \theta$) of the preferred Gabor kernels are displayed in **Figure 4D**. Clearly, the mean of ω has an overall downward trend (Mann–Kendall test, $p < 0.05$) as we go from V1 to LO (the blue line in **Figure 4D**), whereas the mean of σ has an obvious upward trend (Mann–Kendall test, $p < 0.05$) (the green line in **Figure 4D**).

Voxel-to-Voxel Comparisons Between the GaborNet Visual Encoding Model and the Reference Models

Firstly, we compared the encoding performance of the GaborNet-VE model with that of the GWP model, the CNN-linear model, and the ETECR model, respectively. As shown in **Figure 2A**, this is a voxel-to-voxel comparison based on the prediction accuracy of the corresponding voxels. Then, the dominant voxels of each of the two compared models were counted (**Figure 2A**). Finally, we carried out a voxel-to-voxel comparison of the four models and calculated the ratio of the voxels best predicted by each model. In general, the prediction accuracies of the four models for the low-level visual areas (the first row of **Figure 2A**) are higher than those of the intermediate- and high-level visual areas (the second row of **Figure 2A**). The highest prediction accuracy in the low-level visual areas exceeds 0.8, whereas the highest prediction accuracy in the intermediate- and high-level visual areas is only approximately 0.6. Combining **Figures 2A,B**, we demonstrate that the GaborNet-VE model has the strongest overall performance for the low-level visual areas. The percentages of the dominant voxels of the GaborNet-VE model in comparison with each of the three other models are 98, 88, and 83%, respectively (**Figure 2A**). Also, the percentages of the dominant voxels of the GaborNet-VE model in comparison with all of the three other models are 80, 77, and 58% in V1, V2, and V3, respectively. For the intermediate- and high-level visual areas, the GaborNet-VE prediction performance has a strong advantage over the GWP model and is almost equivalent to the performance of each of the other two models. Specifically, the percentages of the

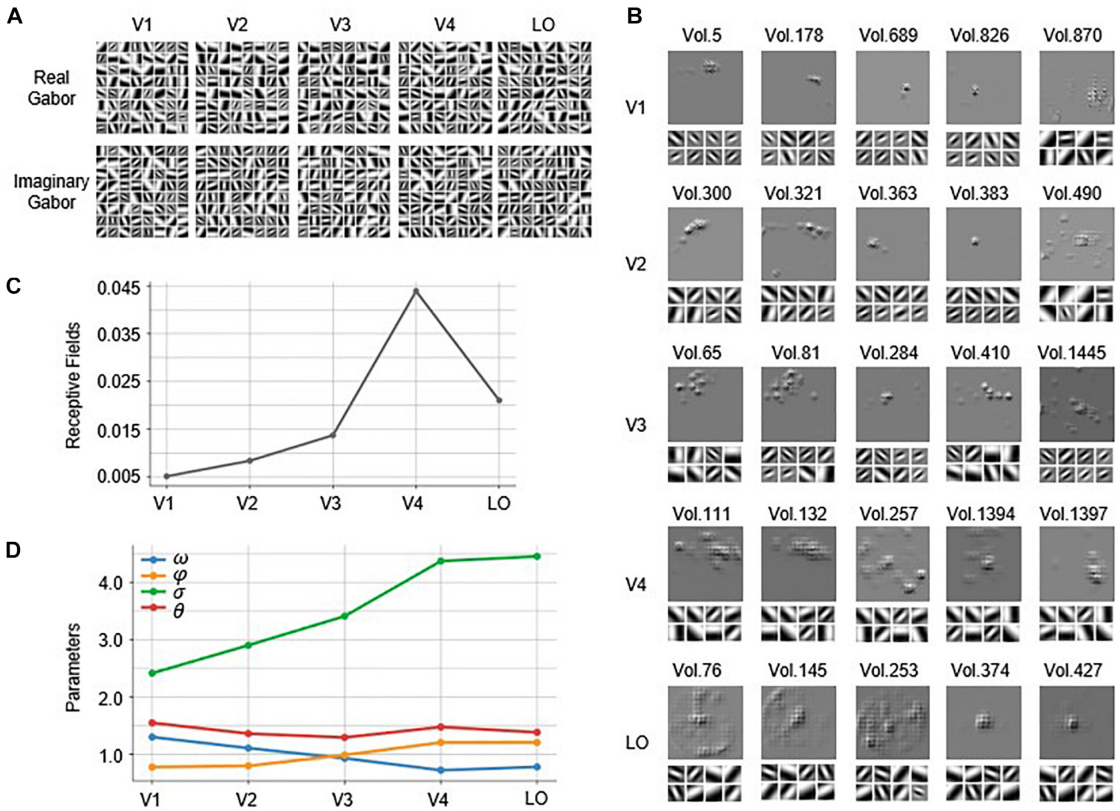


FIGURE 4 | Visualization of the Gabor kernels and RFs. **(A)** Visualization of Gabor kernels. Each of images in first row includes 64 real Gabor kernels learned by real Gabor convolutional layer in GaborNet-VE model for each ROI. Similarly, each of images in second row includes 64 imaginary Gabor kernels learned by imaginary Gabor convolutional layer in GaborNet-VE model for each ROI. **(B)** Visualization of RFs and preferred Gabor kernels. RFs of preferred Gabor kernels for each of five representative voxels are plotted. For each ROI, upper image shows RFs, whereas lower image shows eight preferred Gabor kernels. **(C)** Fitting of RFs to ROIs. Vertical axis represents ratio of RF size to whole image size. Horizontal axis represents five ROIs. **(D)** Tuning of Gabor kernel parameters to ROIs. Vertical axis represents parameter means for preferred Gabor kernels and top 100 best voxel predictions. Blue, orange, green, and red curves represent means of parameters ω, σ, φ, and θ, respectively. Horizontal axis represents five ROIs.

dominant voxels of the GaborNet-VE model in comparison with each of the three other models are 92, 52, and 50%, respectively (**Figure 2A**). As well, the percentages of the dominant voxels of the GaborNet-VE model in comparison with all of the three other models are 37 and 34% in V4, and LO, respectively. Hence, the performance of the GaborNet-VE model has an overall downward trend as we go from V1 to LO. From the perspective of voxel distribution, "banana" shapes are shown in the four diagrams in the first two columns of **Figure 2A**, especially in the upper-middle diagram. These shapes indicate that voxels with high accuracy are mostly on the side of the GaborNet-VE model, whereas voxels with low accuracy are mostly on the side of the reference models. Therefore, voxels with low prediction accuracy under the reference models are effectively rescued by our GaborNet-VE model. Also, these "banana" shapes demonstrate that the GaborNet-VE model provides effective regularization, which benefits from the selective voxel optimization strategy. Because the ETECR model used the same optimization strategy, the voxel distribution of the two diagrams in the last column shows a highly linear pattern (**Figure 2A**).

Model Comparison Based on Voxel Sorting by Prediction Accuracy

Voxels whose responses could be accurately predicted by each of the four models ($\rho > 0.27$) were selected and then sorted in descending order of the prediction accuracy (see **Figure 3**). Overall, results of the GWP model (the gray lines in **Figure 3**) are significantly worse than those of the GaborNet-VE model (the red lines in **Figure 3**), the ETECR model (the blue lines in **Figure 3**), and the CNN-linear model (the green lines in **Figure 3**). Therefore, we focus on comparing the GaborNet-VE model only with the CNN-linear model and the ETECR model. The numbers of voxels in V1, V2, V3, V4, and LO, which are accurately predicted by the GaborNet-VE model, are 548, 688, 336, 223, and 108, respectively. These numbers correspond to voxel percentages of 42.35, 33.03, 18.77, 14.53, and 11.64%, respectively. In the first row of **Figure 3**, the red line is absolutely above the green and blue lines. This demonstrates that the GaborNet-VE model is definitely better than the ETECR and CNN-linear models in predicting the responses in the primary cortex (V1, V2, and V3). However, in the left figure in the second row, the red, green, and blue lines are intertwined. This indicates

that the prediction performance outcomes of the three models for the intermediate visual area (V4) are almost equivalent. In the right figure in the second line, the green line is slightly above the red and blue lines. This indicates a marginal advantage of the CNN-linear model for the LO area. However, regardless of the model selection, the numbers of voxels accurately predicted in V4 and LO are much smaller than those in V1, V2, and V3.

DISCUSSION

We have introduced the GaborNet-VE model, a new end-to-end region-based encoding model for the brain visual ventral stream. The key design principle of this model is constructing a more effective non-linear feature space to fit the brain BOLD responses through learning deep features adapted from handcrafted features. This design was realized by replacing regular convolutional kernels with Gabor kernels with learnable parameters in the first convolutional layer. This combination of the handcrafted and deep features was used to improve the GaborNet-VE model in terms of both expressiveness and biological interpretability, which are viewed as the two most important performance goals of visual encoding models. As the results of this work demonstrated, the lightweight encoding model, composed only of three convolutional layers and one fully connected layer, achieved better prediction accuracy than other compared models for most of the voxels in the visual ventral stream. Moreover, the visualization results showed that the GaborNet-VE model could precisely depict the RF of each voxel and reveal the ROI regularity of fitting to visual features.

Relationship to Previous Work

Although the GaborNet-VE model is an end-to-end model, it can be mapped into the two steps of linearizing encoding models (Naselaris et al., 2011). The three convolutional layers of GaborNet-VE with the ReLU activation function represent the non-linear feature mapping from the input space to the feature space, whereas the fully connected layer represents the linear mapping from the feature space to the activity space. Therefore, the GaborNet-VE model is a special case of the linearizing encoding models, which fits the non-linear feature mapping and the linear mapping together. Several important models have preceded the GaborNet-VE model. One classical model is the GWP model (Kay et al., 2008), which broke through the bottleneck of visual encoding. However, this encoding model is based on handcrafted features, has limitations on feature variations, and hence has poor expressiveness and universality. The GaborNet-VE model could be viewed as an improved variant of the GWP model, where the fixed-parameter Gabor filters were replaced by Gabor filters with learnable parameters, and the non-linear energy mapping was replaced by two convolutional layers with ReLU activation functions. Another widely known important class of models is that of the CNN-linear models (Güçlü and van Gerven, 2015; Eickenberg et al., 2016), which can obtain abundant features to represent voxel responses. However, many unsuitable features were introduced due to the data–task mismatch of this task-driven approach. Moreover, the models get easily trapped into local optima because the models still have two independent steps. The ETECR model addressed the same type of visual encoding problems (Qiao et al., 2019), but it lacks sufficient biological interpretability. The GaborNet-VE model evolved from the ETECR model, whose first convolutional layer was replaced by a parameterized Gabor convolutional layer. The structured features obtained by Gabor kernels improved the interpretability of the end-to-end model.

Advantages of the GaborNet Visual Encoding Model

The GaborNet-VE model presented here follows an effective visual encoding methodology in which an adequate non-linear feature space can be constructed by an end-to-end training process based on the statistics of the visual input and the brain BOLD activities. This neurophysiologic and computational approach accumulates several earlier enhancements and has many advantages in encoding the brain ventral visual pathway. Firstly, the GaborNet-VE model has good expressiveness. The results of this work showed that the encoding model had achieved the state-of-the-art prediction performance for the primary visual cortex, and it is also of comparable prediction performance for the intermediate- and high-level cortex areas. We concluded that the high prediction accuracy could be ascribed to the following reasons: (a) the non-linear feature space learned based on the Gabor features represents the voxel responses in low-level areas; (b) the end-to-end training scheme boosted the performance of the proposed model through learning all optimal parameters simultaneously; (c) the selective optimization strategies for features and voxels led to highly effective features and voxels. Secondly, the GaborNet-VE model has good biological interpretability. The encoding model could learn the relatively precise voxel RFs, rather than roughly measuring the RF size by the feature-weighted RF approach (St-Yves and Naselaris, 2018). As the results suggest, voxels in early visual areas have smaller RFs, whereas voxels further along the ventral pathway have larger RFs. This is consistent with some early findings (Hubel and Wiesel, 1962; Gross et al., 1971; Naselaris et al., 2011). Moreover, the model could learn for each voxel the preferred Gabor kernels, whose properties demonstrated the ROI regularity of tuning to visual features. Overall, the voxels in V1 and V2 prefer the Gabor kernels with high spatial frequencies, whereas the voxels in V4 and LO prefer the Gabor kernels with low spatial frequencies, although few voxels may have opposing preferences in V1, V2, V4, and LO. This phenomenon is consistent with the results of one electrophysiology study, which surprisingly found neurons with low spatial frequencies in V1 and neurons with high spatial frequencies in V4 (Lu et al., 2018). Thirdly, the GaborNet-VE model is highly efficient. The previous voxel-wise encoding models fit one linear regression model for each voxel. Eventually, thousands of encoding models need to be fit for several ROIs. However, the GaborNet-VE model uses an ROI-wise scheme, where all voxels of one ROI are jointly encoded. Therefore, the GaborNet-VE model acts as a single model that needs to be trained for an ROI. Fourthly, the

GaborNet-VE model is a lightweight model. This model merely consists of three convolutional layers and one fully connected layer, whose total number of layers is four. Generally, CNN-based visual encoding modeling approaches can be categorized into either task- or data-driven approaches. In terms of task-driven approaches, AlexNet with eight layers is the most lightweighted DNN. Also, the parameter number of AlexNet is approximately 60M; however, that of GaborNet-VE is approximately 10K. In terms of data-driven approaches, the ETECR model is the most lightweighted DNN, which has five layers. Moreover, we replaced regular convolutional kernels with Gabor convolutional kernels in the first layer, further reduced the number of parameters to be learned. Fifthly, the GaborNet-VE model is relatively robust. It is generally accepted that the network robustness will improve after the regular convolutional kernels are replaced by the structured Gabor kernels.

Limitations and Future Directions

Although the GaborNet-VE model has multiple key advantages, this model still has several disadvantages. The results suggest that the prediction performance of that model for the intermediate and high-level areas is still unsatisfactory (see **Figures 2**, **3**). We noticed as well that the GaborNet-VE model is easily prone to overfitting during training, although the model is lightweight. On the other hand, voxels in high-level visual areas depend on a more complex feature representation, which requires a network with more layers. Therefore, a large fMRI dataset is crucially needed. A common problem of current visual encoding models is that they cannot accurately express neural activities in higher visual areas due to the limited fMRI samples. To weaken the influence of data scarcity, we can implement advanced few-shot deep learning and meta-learning techniques, such as the meta-learning approach. The GaborNet-VE model has both good expressiveness and biological interpretability, and this validates the advantages of combining deep and handcrafted features. Therefore, a future research direction should address how to combine deep network features with handcrafted ones more effectively.

CONCLUSION

We have proposed the GaborNet-VE model, a new end-to-end ROI-wise encoding model for the brain ventral visual stream. The results of this study demonstrated that the GaborNet-VE model achieved state-of-the-art prediction performance for the primary visual cortex. Hence, the deep-network feature space learning based on Gabor features is more expressive and effective. Moreover, the GaborNet-VE model has good biological interpretability, which could be demonstrated by the ROI regularity of tuning to visual features and the visualization of the estimated RFs. Hence, the proposed lightweight visual encoding model has both good expressiveness and biological interpretability based on combining deep and handcrafted features.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://crcns.org/data-sets/vc/vim-1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the UC Berkeley Office for Protection of Human Subjects (cphs.berkeley.edu). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YC, KQ, and LW performed the conceptualization and investigation. YC and KQ performed the data curation, formal analysis, and methodology. YC, KQ, and CZ performed the software. CZ, LW, BY, and LT performed the validation. YC, KQ, CZ, LW, and BY performed the visualization and writing – original draft. LT provided article review and assistance with the final article write up. BY and LT performed the funding acquisition, supervision, and project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Key Research and Development Plan of China (No. 2017YFB1002502).

REFERENCES

Adelson, E. H., and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Optical Soc. Am. A* 2, 284–299. doi: 10.1364/JOSAA.2.000284

Agrawal, P., Stansbury, D., Malik, J., and Gallant, J. L. (2014). Pixels to voxels: modeling visual representation in the human brain. *arXiv* [Preprint]. doi: arxiv.org/abs/1407.5104

Alekseev, A., and Bobe, A. (2019). GaborNet: gabor filters with learnable parameters in deep convolutional neural networks. *arXiv* [Preprint]. doi: arxiv.org/abs/1904.13204

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., et al. (2019). Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLoS Comput. Biol.* 15:e1006897. doi: 10.1371/journal.pcbi.1006897

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., et al. (2005). Do we know what the early visual system does? *J. Neurosci.* 25, 10577–10597. doi: 10.1523/JNEUROSCI.3726-05.2005

Cichy, R. M., and Kaiser, D. (2019). Deep neural networks as scientific models. *Trends Cogn. Sci.* 23, 305–317. doi: 10.1016/j.tics.2019.01.009

Eickenberg, M., Gramfort, A., Varoquaux, G., and Thirion, B. (2016). Seeing it all: convolutional network layers map the function of the human visual system. *NeuroImage* 152, 184–194. doi: 10.1016/j.neuroimage.2016.10.001

Gabor, D. (1946). Theory of communications. part 1: the analysis of information. *J. Institution Electr. Eng.* 93, 429–457. doi: 10.1049/ji-3-2.1946.0074

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1971). Visual properties of neurons in inferotemporal cortex of the Macaque. *J. Neurophysiol.* 35, 96–111. doi: 10.1152/jn.1972.35.1.96

Güçlü, U., and van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* 35, 10005–10014. doi: 10.1523/JNEUROSCI.5023-14.2015

Güçlü, U., and van Gerven, M. A. J. (2017). Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *NeuroImage* 145, 329–336. doi: 10.1016/j.neuroimage.2015.12.036

Han, K., Wen, H., Shi, J., Lu, K.-H., Zhang, Y., and Liu, Z. (2017). Variational autoencoder: an unsupervised model for modeling and decoding fMRI activity in visual cortex. *NeuroImage* 198, 125–136. doi: 10.1016/j.neuroimage.2019.05.039

Huang, L.-L., Shimizu, A., and Kobatake, H. (2004). *Classification-Based Face Detection Using Gabor Filter Features*. New York, NY: IEEE.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243. doi: 10.1113/jphysiol.1968.sp008455

Hung, C. P., Kreiman, G., Poggio, T., and Dicarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866. doi: 10.1126/science.1117593

Huth, A. G., Nishimoto, S., Vu, A. T., and Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76, 1210–1224. doi: 10.1016/j.neuron.2012.10.014

Jones, J. P., and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58, 1233–1258. doi: 10.1152/jn.1987.58.6.1233

Kay, K. N. (2018). Principles for models of neural information processing. *NeuroImage* 180, 101–109. doi: 10.1016/j.neuroimage.2017.08.016

Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature* 452, 352–355. doi: 10.1038/nature06713

Kietzmann, T. C., Mcclure, P., and Kriegeskorte, N. (2017). Deep neural networks in computational neuroscience. *bioRxiv* [Preprint]. doi: 10.1101/133504

Kingma, D., and Ba, J. (2014). "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, (San Diego).

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Adv. Neural. Inform. Process. Syst.* 60, 84–90. doi: 10.1145/3065386

Kyrki, V., Kamarainen, J.-K., and Kälviäinen, H. (2004). Simple gabor feature space for invariant object recognition. *Pattern Recognit. Lett.* 25, 311–318. doi: 10.1016/j.patrec.2003.10.008

Li, C., Xu, J., and Liu, B. (2018). Decoding natural images from evoked brain activities using encoding models with invertible mapping. *Neural Netw.* 105, 227–235. doi: 10.1016/j.neunet.2018.05.010

Liu, C., and Wechsler, H. (2001). *A Gabor Feature Classifier for Face Recognition*. New York, NY: IEEE.

Liu, C., and Wechsler, H. (2002). *Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition*. New York, NY: IEEE.

Lu, Y., Yin, J., Chen, Z., Gong, H., Liu, Y., Qian, L., et al. (2018). Revealing detail along the visual hierarchy: neural clustering preserves acuity from V1 to V4. *Neuron* 98, 417–428.e3. doi: 10.1016/j.neuron.2018.03.009

Meshgini, S., Aghagolzadeh, A., and Seyedarabi, H. J. (2012). Face recognition using gabor filter bank, kernel principle component analysis and support vector machine. *Int. J. Comput. Theory Eng.* 4, 767–771. doi: 10.7763/IJCTE.2012.V4.574

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science* 320, 1191–1195. doi: 10.1126/science.1152876

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, (Haifa: Omnipress).

Naselaris, T., Kay, K. N., Nishimoto, S., and Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage* 56, 400–410. doi: 10.1016/j.neuroimage.2010.07.073

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., and Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron* 63, 902–915. doi: 10.1016/j.neuron.2009.09.006

Naselaris, T., Stansbury, D. E., and Gallant, J. L. (2012). Cortical representation of animate and inanimate objects in complex natural scenes. *J. Physiol. Paris* 106, 239–249. doi: 10.1016/j.jphysparis.2012.02.001

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646. doi: 10.1016/j.cub.2011.08.031

Oquab, M., Bottou, L., Laptev, I., and Sivic, J. (2014). *Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks*. New York, NY: IEEE. doi: 10.1109/CVPR.2014.222

Qiao, K., Zhang, C., Chen, J., Wang, L., Tong, L., and Yan, B. (2019). Effective and efficient ROI-wise visual encoding using an end-to-end CNN regression model and selective optimization. *arXiv* [Preprint]. doi: arXiv:1907.11885

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., et al. (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770. doi: 10.1038/s41593-019-0520-2

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. J. N. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536. doi: 10.1038/323533a0

Serre, T. (2019). Deep learning: the good, the bad, and the ugly. *Annu. Rev. Vis. Sci.* 5, 399–426. doi: 10.1146/annurev-vision-091718-014951

Shi, J., Wen, H., Zhang, Y., Han, K., and Liu, Z. (2018). Deep recurrent neural network reveals a hierarchy of process memory during dynamic natural vision. *Hum. Brain Mapp.* 39, 2269–2282. doi: 10.1002/hbm.24006

Springenberg, J., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for simplicity: the all convolutional net. *arXiv* [Preprint]. doi: arXiv:1412.6806

Stansbury, D. E., Naselaris, T., and Gallant, J. L. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron* 79, 1025–1034. doi: 10.1016/j.neuron.2013.06.034

St-Yves, G., and Naselaris, T. (2018). The feature-weighted receptive field: an interpretable encoding model for complex feature spaces. *NeuroImage* 180, 188–202. doi: 10.1016/j.neuroimage.2017.06.035

Ukita, J., Yoshida, T., and Ohki, K. (2019). Characterisation of nonlinear receptive fields of visual neurons by convolutional neural network. *Sci. Rep.* 9:3791. doi: 10.1038/s41598-019-40535-4

Vu, V. Q., Ravikumar, P., Naselaris, T., Kay, K. N., Gallant, J. L., and Yu, B. (2011). Encoding and decoding V1 fMRI responses to natural images with sparse nonparametric models. *Ann. Appl. Stat.* 5, 1159–1182. doi: 10.1214/11-AOAS476

Wen, H., Shi, J., Chen, W., and Liu, Z. (2018). Deep residual network predicts cortical representation and organization of visual features for rapid categorization. *Sci. Rep.* 8:3752. doi: 10.1038/s41598-018-22160-9

Yamins, D. L., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244

Zeiler, M. D., and Fergus, R. (2013). Visualizing and understanding convolutional networks. *arXiv* [Preprint]. doi: arXiv:1311.2901

Zhang, C., Qiao, K., Wang, L., Tong, L., Hu, G., Zhang, R.-Y., et al. (2019). A visual encoding model based on deep neural networks and transfer learning for brain activity measured by functional magnetic resonance imaging. *J. Neurosci. Methods* 325:108318. doi: 10.1016/j.jneumeth.2019.108318

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 10 February 2021
doi: 10.3389/fnins.2020.618481

Artificial Intelligence Segmented Dynamic Video Images for Continuity Analysis in the Detection of Severe Cardiovascular Disease

Xi Zhu[1†], Wei Xia[1†], Zhuqing Bao[1], Yaohui Zhong[2], Yu Fang[1], Fei Yang[1], Xiaohua Gu[1], Jing Ye[1] and Wennuo Huang[1]**

[1] Clinical Medical College, Yangzhou University, Yangzhou, China, [2] Department of Computer Science and Technology, Nanjing University, Nanjing, China

In this paper, an artificial intelligence segmented dynamic video image based on the process of intensive cardiovascular and cerebrovascular disease monitoring is deeply investigated, and a sparse automatic coding deep neural network with a four layers stack structure is designed to automatically extract the deep features of the segmented dynamic video image shot, and six categories of normal, atrial premature, ventricular premature, right bundle branch block, left bundle branch block, and pacing are achieved through hierarchical training and optimization. Accurate recognition of heartbeats with an average accuracy of 99.5%. It provides technical assistance for the intelligent prediction of high-risk cardiovascular diseases like ventricular fibrillation. An intelligent prediction algorithm for sudden cardiac death based on the echolocation network was proposed. By designing an echolocation network with a multilayer serial structure, an intelligent distinction between sudden cardiac death signal and non-sudden death signal was realized, and the signal was predicted 5 min before sudden death occurred, with an average prediction accuracy of 94.32%. Using the self-learning capability of stack sparse auto-coding network, a large amount of label-free data is designed to train the stack sparse auto-coding deep neural network to automatically extract deep representations of plaque features. A small amount of labeled data then introduced to micro-train the entire network. Through the automatic analysis of the fiber cap thickness in the plaques, the automatic identification of thin fiber cap-like vulnerable plaques was achieved, and the average overlap of vulnerable regions reached 87%. The overall time for the automatic plaque and vulnerable plaque recognition algorithm was 0.54 s. It provides theoretical support for accurate diagnosis and endogenous analysis of high-risk cardiovascular diseases.

Keywords: artificial intelligence, segmented dynamic, video imaging, detection of severe cardiovascular disease, continuity analysis

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing Medical University, China
Yufeng Yao,
Changshu Institute of Technology, China

***Correspondence:**
Jing Ye
yhzejing@163.com
Wennuo Huang
huangwennuo@sina.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *17 October 2020*
Accepted: *11 November 2020*
Published: *10 February 2021*

Citation:
Zhu X, Xia W, Bao Z, Zhong Y, Fang Y, Yang F, Gu X, Ye J and Huang W (2021) Artificial Intelligence Segmented Dynamic Video Images for Continuity Analysis in the Detection of Severe Cardiovascular Disease. Front. Neurosci. 14:618481. doi: 10.3389/fnins.2020.618481

INTRODUCTION

Chronic non-communicable diseases (NCDs) have become a major public health problem affecting the country's economic and social development, with mortality accounting for 86.6% of the entire spectrum of diseases, and the burden of the disease accounting for more than 70% (Luo et al., 2018). Among the hazards of chronic diseases, cerebrovascular, and cardiovascular diseases (CCVD)

have been occupying the first place in the whole spectrum of diseases in terms of morbidity and mortality, and the prevalence rate is on the rise (Bechar et al., 2018). Electrocardiogram (ECG) and medical imaging are the main tools for clinical diagnosis of cardiovascular disease. ECG includes a short-time electrocardiogram (EKG) collected at rest and a long-time dynamic electrocardiogram (Holter) collected under moving conditions (Sonkusare et al., 2019). The main advantages of ECG are that it is easy, inexpensive, and portable to collect electrical signals from outside the body and that it responds to the electrophysiological activity of the heart (Nascimento and Carneiro, 2017). Compared with electrocardiogram, it can reveal the internal activity of the heart. Its disadvantages are the fixed collection position, long collection time, high price, immobility, and the inability to realize real-time collection (Latha et al., 2020). Cardiovascular disease has both covert and sudden features. The covert nature of the disease determines that ECG cannot be observed only, and more accurate medical images are needed to discover hidden diseases and symptoms, to carry out timely intervention and treatment; the sudden nature of the disease also determines that medical images cannot be used only, and ECG needs to be portable to monitor the condition of the heart in real-time and provide timely warning of possible dangers (Randive et al., 2020). Only when ECG and medical imaging cooperate, can the mortality rate of cardiovascular disease be effectively reduced (Musial et al., 2020). However, the existing dynamic ECG equipment still adopts offline storage mode, which cannot realize timely signal analysis and early warning; medical image analysis is limited by the professional level of doctors themselves, and at the same time is easily interfered by other factors, which can easily produce inconsistent results among different doctors, and cannot achieve the purpose of accurate diagnosis and treatment (Aminikhanghahi and Cook, 2017). Therefore, the research on automatic analysis algorithms of ECG and medical images based on artificial intelligence plays a crucial role in the prevention and treatment of cardiovascular diseases (Darwish et al., 2020).

In arrhythmia analysis, Costa et al. (2019) proposed to extract arrhythmia-related features from the double-tree complex wavelet transform coefficients and classify them by multilayer neural network, which has higher sensitivity than discrete wavelet transform coefficients. Shinbane and Saxon (2018) proposed to extract RR interval features, higher-order statistical features, and Gaussian mixed model parameters from ECG signals, which is more sensitive than discrete wavelet transform coefficients. Characteristics to classify arrhythmias, and by training the decision tree model, arrhythmias were identified with an accuracy of 99.7%. In the diagnosis of premature ventricular beats, Xiao et al. (2019) proposed a fractional linear prediction method for the detection of premature ventricular beats and demonstrated that it has higher accuracy than linear prediction and a higher sensitivity to premature ventricular beats than other beats. Ventricular premature heartbeat, establish monitoring statistical model by discrete wavelet decomposition, and set the upper limit value of the model, when the monitoring statistical model parameters exceed the upper limit value, to alarm (Haider et al., 2019). When the parameters of the monitoring statistical model exceed the upper limit, the alarm will be raised (Fan et al., 2019). The accuracy of premature ventricular beats detection is 97.9% as verified by MIT-BIH arrhythmia database data (Deldari et al., 2020). In the aspect of myocardial infarction prediction, Yang et al. (2020) proposed a convolutional neural network method to identify myocardial infarction heartbeat and normal heartbeat in the ECG signal, and the recognition accuracy reached 93.53 and 95.22% through the noise-free environment and noise-free environment tests, respectively. Saggi and Jain (2018) also applied the convolutional neural network to automatically identify calcification in the coronary artery in coronary CT images. With a plaque recognition accuracy of 85% for images with cardiac motion disturbances (Saggi and Jain, 2018). However, due to the limitations of CT image acquisition and its resolution, it is not able to accurately analyze the intracoronary plaque composition, hence the emergence of endovascular medical images such as VIES and OCT (Hong et al., 2018). In contrast, Craye et al. (2016) extracted 54 sets of features from VIES images to describe fibrotic lipids, calcifications, necrotic cores, and fibrotic plaques in blood vessels and performed feature screening by PCA algorithm to achieve accurate classification of intravascular plaques. Different migratory learning methods used to establish different forms of image expressions, and eventually, the accurate detection of plaque tissue was achieved by fusing multiple representations, with a model detection accuracy of 91.7% (Craye et al., 2016). Pasterkamp also extracted a series of plaque-related geometric and non-geometric features by taking advantage of the high-resolution characteristics of coronary OCT images, and then applied the SVM classifier to classify the relevant features to achieve accurate identification of fibrotic, lipid, and calcified plaques, with an average identification accuracy of 94.0, 97.2, and 99.2%, respectively (Pasterkamp, 2018).

The main task of video image behavior recognition is to analyze a video image and then classify the human behaviors contained in it. The difficulty in video image behavior recognition comes from two aspects: first, the time variability, there may be gaps between the actions that have nothing to do with the behavior, the time point of the behavioral actions is uncertain, and the continuous interval between the actions is also different. Therefore, given a video image to be able to identify the start and end time of the behavior, for the video image frames that have nothing to do with the behavior, weaken its role or discard. Second, the spatial complexity, different perspectives, illumination, and background will cause different scenes, different scenes of the same behavior action will produce certain differences. Even in fixed scenes, behavioral actions can vary depending on the perspective of the person, individual differences, and shading. These can have an impact on accuracy. Current algorithms are dedicated to solving the problem of how to extract better features that describe the judgments made in video images and better temporal information in video images. Mainstream methods are using partial continuous frames of video images and long duration video image information, which results in a lot of missing information and there is a lot of redundant information in video images, and randomly selecting video image frames may miss a lot of important information. So, in this paper, the attention mechanism is used to empower the video image frames to weaken the redundant information and

increase the useful video image frames to influence the result. In medical imaging, the coronary vascular morphology assessment and plaque identification algorithms closely related to high-risk cardiovascular diseases were investigated. The automatic endovascular contour extraction algorithm in coronary OCT images was investigated to achieve accurate extraction and 3D modeling of coronary vessels, and the adaptiveness of the coronary endovascular extraction algorithm to different feature images was realized through the grayscale distribution analysis of OCT images, which supported the accurate assessment of coronary vessel morphology.

ARTIFICIAL INTELLIGENCE SEGMENTED DYNAMIC VIDEO IMAGES IN CONTINUITY ANALYSIS DESIGN

Artificial Intelligence Modeling

In terms of the composition of the entire autoencoder network, the implicit layer can be viewed as a representation of features extracted from the input layer, and then the input is reproduced from these features by way of reverse coding (Banerjee et al., 2019). Therefore, the dimensionality of the implicit layer is usually smaller than the dimensionality of the input layer. The process of training the network is the process of continually reducing the error between the input and output, so the loss function of the network can be defined as:

$$K(W,b) = \frac{1}{2p}\sum_{i=1}^{p}\left|\left|X^i - Y^i\right|\right|_2^2 + \frac{\lambda}{2}\left(\left|\left|W_e\right|\right|_2^2 + \left|\left|W_d\right|\right|_2^2\right) \quad (1)$$

The first term is the mean-variance of the input and output data, and the second term is the attenuation term of the encoding and decoding network weight parameters, which is mainly used to reduce the update rate of the weights and prevent the data training from converging to the local optimum, n is the number of input samples, and mm is the attenuation coefficient of the weights (Diehn et al., 2016).

Different tasks have different feature complexity, so different numbers of implicit layer nodes are often required when applying autoencoder networks to solve practical problems (Eldib et al., 2018). The characteristic of the autoencoder implicit layer dimension is lower than the input data dimension can sometimes limit its application. To solve this problem, sparse rules are added to the autoencoder network. Sparsity is the selective activation of a small number of nodes in the implicit layer nodes, leaving most nodes in a suppressed state (Ali et al., 2018). This allows the number of implicit layer nodes in the network to be greater than the number of input layer nodes, and depending on the complexity of the task, different levels of sparsity can be set to meet the practical requirements, as shown in **Figure 1**.

Thus, an average activation function is introduced into the auto-coding network (Salekin et al., 2018), defined as:

$$\overline{p_j} = \frac{1}{p}\sum_{i=1}^{p} H_i(X^j) \quad (2)$$

To ensure that the activation of the implicit layer nodes is small, a relative entropy function is introduced into the model, which is expressed as follows:

$$PL(p||\overline{p_j}.) = p\log\frac{p}{\overline{p_j}} + (1-p)\log\frac{1-p}{1-\overline{p_j}} \quad (3)$$

The network training process is the process of progressively updating the network weight parameters, propagating to the implicit layer to get tr, X through the formula (3) propagates back to the data layer to get w, w again passed to the implicit layer to get T, then the formula for updating the network weight parameters can be expressed as.

$$tr(w^T X L_N \mathrm{X}^{\mathrm{Tw}}) = tr(w^T X L \mathrm{X}^{\mathrm{Tw}}) \quad (4)$$

After completing the training of one layer of autoencoder, the feature representation in the implicit layer is obtained, and the feature representation obtained in this layer is used as the data input for the next layer, and the next layer of the network is trained to obtain the features of the next implicit layer, and so on to complete the training of the entire stack sparse autoencoder network layer by layer (Horton et al., 2020).

In this paper, a stack sparse autoencoder network with four layers is constructed, and the corresponding nodes from the first to the fourth layer are 120, 60, 30, and 15, respectively. The sparse autoencoder of the first layer is trained by the input heartbeat, and the feature output of the first layer is obtained. Then, the sparse autoencoder of the second layer is trained with the output of the first layer to obtain the second layer's features. Finally, the four layers of sparse autoencoders are trained, and the depth feature of the cardioid signal is obtained. The depth features of the cardioid signal are extracted automatically during the network training, which solves the problem of incomplete features or redundancy between features when they are selected manually.

The error proportion from the feature implicit layer to the data output layer first calculated, which is expressed as:

$$\sigma_i^2 = \left(\left(\sum_{j=1}^{S} W_{ij}^{(2)}\sigma_j^{(3)}\right) + \beta(-p/\overline{p_j} + (1-p)/(1-\overline{p_j})\right) \quad (5)$$

A direct transfer function corresponds to the physiological system of viewing aortic blood pressure as input and radial blood pressure as output, and to be able to reconstruct central arterial pressure from the radial arterial sphygmomanometer signal, an inverse transfer function would be derived from the above equation.

$$P(t-1) = \frac{-b_2 P(t-2)}{b_1} - \ldots \frac{-b_{nb} P(t-nb)}{b_1} \quad (6)$$

Fisher vector is an encoding method that enables the normalization of unequal feature matrices. Existing classification methods fall into two main categories: generative methods, such as GMM (Lee et al., 2019), which reflect the similarity between similar data, and discriminative methods, such as SVM, which reflect the differences between dissimilar data. The two

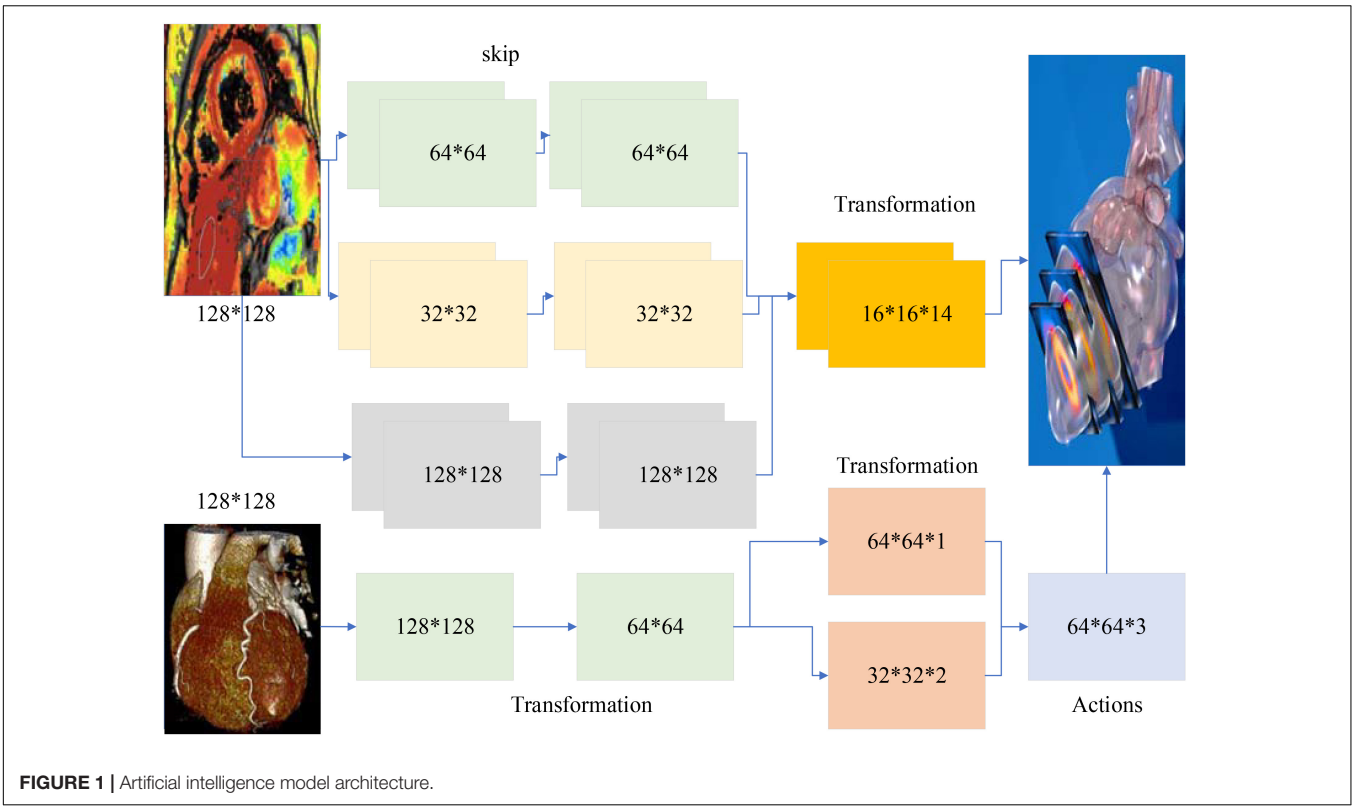


FIGURE 1 | Artificial intelligence model architecture.

advantages that work better will be used to generate models for use in the discriminant classifier.

$$\gamma_t(i) = p(i\,|\lambda_t,\ \lambda) = \frac{w_i.p_i(\lambda_t\,|\lambda)}{\sum_{i=1}^{M} w_i.p_i(\lambda_t\,|\lambda)} \tag{7}$$

A first bias derivation of the parameters yields (Ren et al., 2019):

$$\frac{\partial L(\lambda_t\,|\lambda)}{\partial w_i} = \sum_{i=1}^{T}\left[\frac{\gamma_i(i)}{w_i} - \frac{\gamma_t(1)}{w_i}\right], i > 1 \tag{8}$$

$$\frac{\partial L(\lambda_t\,|\lambda)}{\partial w_i^d} = \sum_{i=1}^{T}\gamma_t^i\left[\frac{\gamma_i(i)^2}{w_i} - \frac{\gamma_t(1)^2}{w_i}\right] \tag{9}$$

$$\frac{\partial L(\lambda_t\,|\lambda)}{\partial w_i^d} = \sum_{i=1}^{T}\gamma_t^i\left[\frac{\gamma_i(i)^2}{w_i} - \frac{1}{w_i}\right] \tag{10}$$

The optical flow field is used to obtain the trajectory information in the video sequence: to sample the feature points densely at multiple scales of the picture, respectively, a partition grid is used to filter out the points with few transformations; to obtain the motion velocity of the feature points, the median optical flow in the neighborhood of the feature points is computed, and then the key points are tracked; a total of four types of HOG, HOF, trajectory, and MBH are extracted along with the trajectory information. The HOG feature is based on the grayscale image calculation, and several other features are based on the dense optical flow field calculation. The HOG feature is the gradient amplitude of a pixel counted according to its gradient direction after blocking. The HOF feature is obtained by calculating the grayscale transformation matrix and gradient matrix of the optical flow adjacent to the current frame at a time, and then weighting the optical flow direction.

The most important improvement is the use of camera motion estimation to remove trajectories and optical flows present in the background. In the DT algorithm, as the camera moves, it generates many trajectories on the background, and this motion information can greatly affect the trajectory of the human body. This trajectory information is noisy and has a weak relationship with the identified behavior. So, there is a need to eliminate this type of noise. The motion of the trajectory is calculated by computing the optical flow information, so to eliminate the background the optical flow can be estimated by estimating the camera movement. Since the variation between two adjacent frames is small, the algorithm assumes that a projection transformation matrix is used to describe the relationship between two adjacent images, i.e., the latter frame is transformed by the projection of the previous frame. Thus, the estimation of camera motion becomes a matter of computing the transformation projection matrix from the images of the previous and previous frames.

Segmented Moving Video Images in Continuity Analysis Design

The classical dual-stream network model is divided into a temporal network, where the input to the temporal network is

a stack of optical frames, and a spatial network, where the input to the spatial network is a single RGB image (Humphreys and Wang, 2018). This approach can only use a limited number of video frames in the video, but not the entire video information, besides, the spatial network input is a single random RGB image, which can only capture the appearance of a frame, there are some frames in the video that has nothing to do with the action, so it is a big mistake to extract appearance information from a random frame. The time correlation between successive frames is extracted by superimposed optical flow frames. For the convolutional network, if the number of convolutional kernels is much smaller than the number of superimposed optical flow frames, the information between frames is lost. The overall block diagram of the network is shown in **Figure 2**.

The input of the whole network is a long duration image sequence containing a motion, which is divided into several overlapping segments by video pre-processing. For each overlapping segment, it includes the RGB frame segment and optical stream frame segment, respectively, and the RGB frame segment is used as input to the spatial network; while the optical stream frame is a single channel, to keep the three-channel input of each sub-network consistent, the optical stream u, optical stream v and the mean value between them are used to form a three-channel picture, and the continuous optical stream frame is used as input to the temporal network (Mariakakis et al., 2017). The whole network is divided into several branches, and the layer number and hyper-parameter settings of each sub-network are the same before feature fusion, which can facilitate the training of the network, and also ensure that the features are consistent in dimension when fusing different stream features, which provides convenience for the fusion of different network features. The design idea of the network structure proposed in this paper is divided into five modules, the first module is the video pre-processing module, the second module is the convolutional neural network plus LSTM extraction of spatiotemporal features module, the third module is the attention mechanism module, the fourth module is the long duration information fusion module, and the fifth module is the classification module. The second and third module is called attentional mechanism feature extraction network, video frame by frame through the convolutional network to extract spatial features, through the LSTM decoding, adding attention mechanism in the decoding process to learn the weight of the impact of each frame on the results, and finally through the SoftMax layer output prediction results.

Video pre-processing module, the main function of this module is to process the original video and divide the input video into several clips. One-third of the length of each clip overlaps each other to maintain the continuity of the information. The input of the network is divided into RGB raw frame and optical stream frame, and the video is transformed into the corresponding picture to be saved by the coding and decoding method and optical stream algorithm in OpenCV.

Video pre-processing consists of two parts: segmentation and optical stream extraction. To extract the temporal and spatial features of the long-term video, we divide the input video into Sn fragments. One-third of the length of each fragment overlaps each

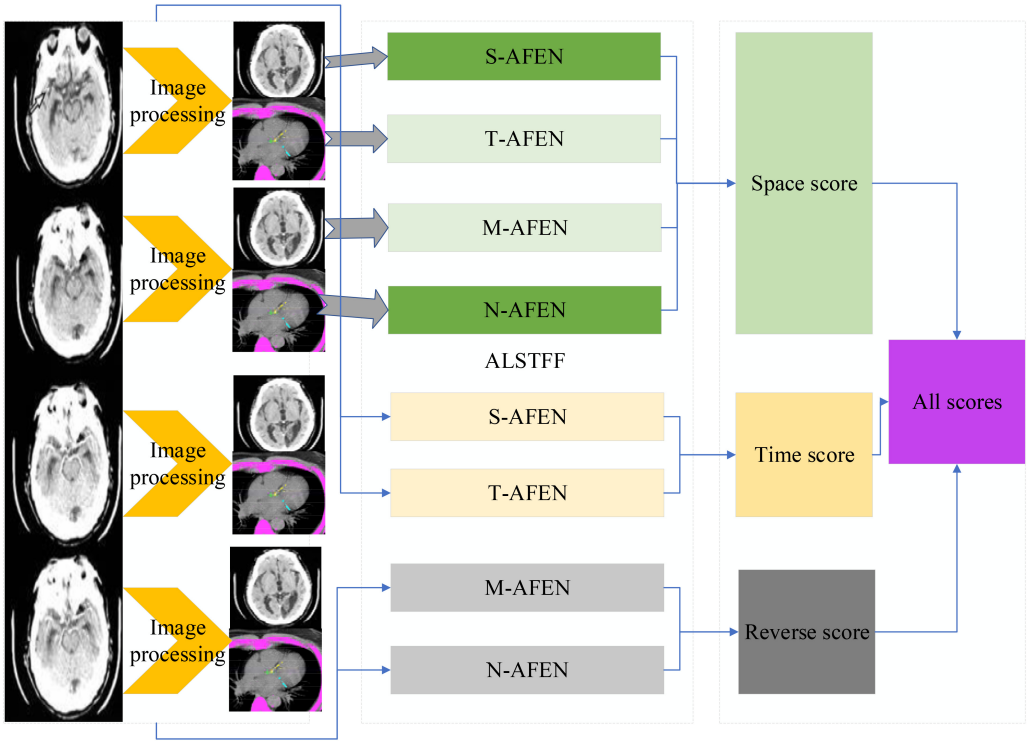


FIGURE 2 | ALSTFF overall flow chart.

other to maintain the continuity of the information. The input of the whole network is divided into two major pieces, RGB video frames, and optical stream frames (Jeannot, 2019). RGB frames can be read directly from the video in OpenCV and each frame of the video is saved as a picture. The optical stream frame is calculated by an algorithm that extracts the optical stream. In this paper, the TV-L1 algorithm is selected as the optical stream extraction algorithm.

Due to the relatively small amount of data in the two datasets, using a small sample of data to train the network can easily lead to overfitting of the network. The current state-of-the-art neural networks require 1,000 of images to train the network to perform well (Eriksson and Eriksson, 2019). If there is not a lot of data, data augmentation is one of the ways to increase the amount of data, we do not have to look for novel images to add to the data set, because in neural network training, the recognition of images is not so intelligent, and inadequately trained neural network will think that the same object at different positions and angles belong to different objects, while in the eyes of humans, the object is only by displacement and change the angle. So, to get more data, just make small changes to the existing dataset, like rotation, displacement, flipping, etc. Our network will treat these images as different. Video belongs to a sequence problem which includes both spatial and temporal sequences, to solve such a problem we need to use RNN, the structure of RNN is shown schematically in **Figure 3**.

Video behavior recognition currently exists in algorithms, both manual feature-based, and deep learning-based methods, to extract spatial and temporal information from the video, spatial information contains information about the appearance of the video, and temporal information contains information about the temporal continuity between frames of the video. The spatial information is usually extracted using a convolutional neural network, and the temporal information is usually extracted using a cyclic neural network, which uses different filters to get different picture features. Different convolutional kernels correspond to different picture effects. The convolutional neural network can learn the features of the image in a supervised way, that is, it can learn the values inside the convolutional kernel, and through the backpropagation mechanism, the error is transmitted back, and the values in the convolutional kernel are updated at each layer by automatic derivation, which is the so-called gradient, and finally the error is minimized. At the same time, after each layer of convolution, all need to go through a non-linear activation function, adding non-linear factors in the neural network, so that the network can learn more complex functional characteristics, if you do not use non-linear activation function, then the output of the model is just a linear combination of the input, even if there are multiple hidden layers, if you use a linear activation function or no activation function, then the network has always done is linear. So, it is equivalent to having no hidden layer, the neural network just linearly combines the inputs and then outputs. The pooling operation is performed after the convolutional layer, which on the one hand reduces the dimensionality of the data, and at the same time reduces the redundant features in the image. Thus, theoretically the deeper the network, the richer the features that can be learned and the better the results.

ANALYSIS OF THE PROCESS OF VIDEO IMAGE SURVEILLANCE DETECTION OF SEVERE CARDIOVASCULAR DISEASE

Design of Video Imaging Tests for Critical Cardiovascular Disease

Central arterial pressure measured by an invasive/minimally invasive device is used as the gold standard reference, and

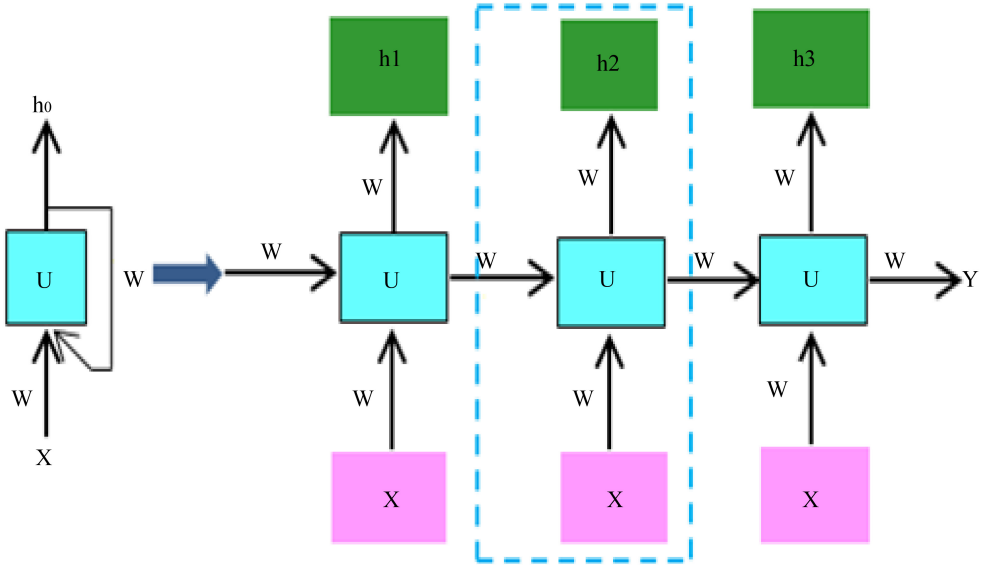


FIGURE 3 | RNN structure diagram.

peripheral blood pressure measured by a non-invasive finger sphygmomanometer is reconstructed by an algorithm to evaluate the accuracy of the algorithm by comparing the central arterial pressure with its analysis error. The number of stable patients requiring interventional intubation (the head of the interventional guidewire is placed at 8 cm from the heart valve, or close to it) is more than or equal to 10. Experimental preparation phase: the operator prepares the experiment, including putting on clothing, gloves, and mask, placing the device in the appropriate position, and connecting the interface cable; the subject lies still on the experimental bed for 5 min; the subject wears the sensor component on the left finger; the continuous blood pressure measurement device is started until the finger blood pressure data can be read correctly, and The interventionalist inserts a micro-manometer catheter equipped with a special blood pressure measuring device into the aortic root by puncturing the femoral/radial/brachial artery and places the head of the catheter about 8 cm away from the heart valve, and the external sensor is collected by a data acquisition card and transferred to the computer for adjusting the blood pressure waveform at the central artery. After the waveform readings are stable and the absolute values are accurate, record the waveform data for 1 min; withdraw the probe and remove the peripheral blood pressure measurement equipment to complete the data acquisition for one patient, and repeat the above experimental steps for the other patients, as shown in **Figure 4**.

TABLE 1 | Statistics of experimental results.

Patient number	RMSE (mmHg)	PTT (ms)
1	3.4	34
2	4.1	24
3	4.7	75
4	5.7	77
5	5.9	83
6	5.5	42
7	3.5	72
Average value	2.9	71
Variance	5.8	60

A total of seven minimally invasive surgery patients were collected, and their basic conditions are shown in **Table 1**. The time delay PTT of the blood pressure wave from the finger to the heart was obtained using the estimation method. This study was approved by the ethics committee and the patients provided consent forms.

The procedure consisted of each subject has his or her head slightly elevated and breathing steadily in a supine position for approximately 5 min before the start of the experiment. Continuous finger blood pressure was then monitored by a non-invasive blood pressure device at the same level as the heart for 5 min or more to ensure that the smooth muscle of the fingers

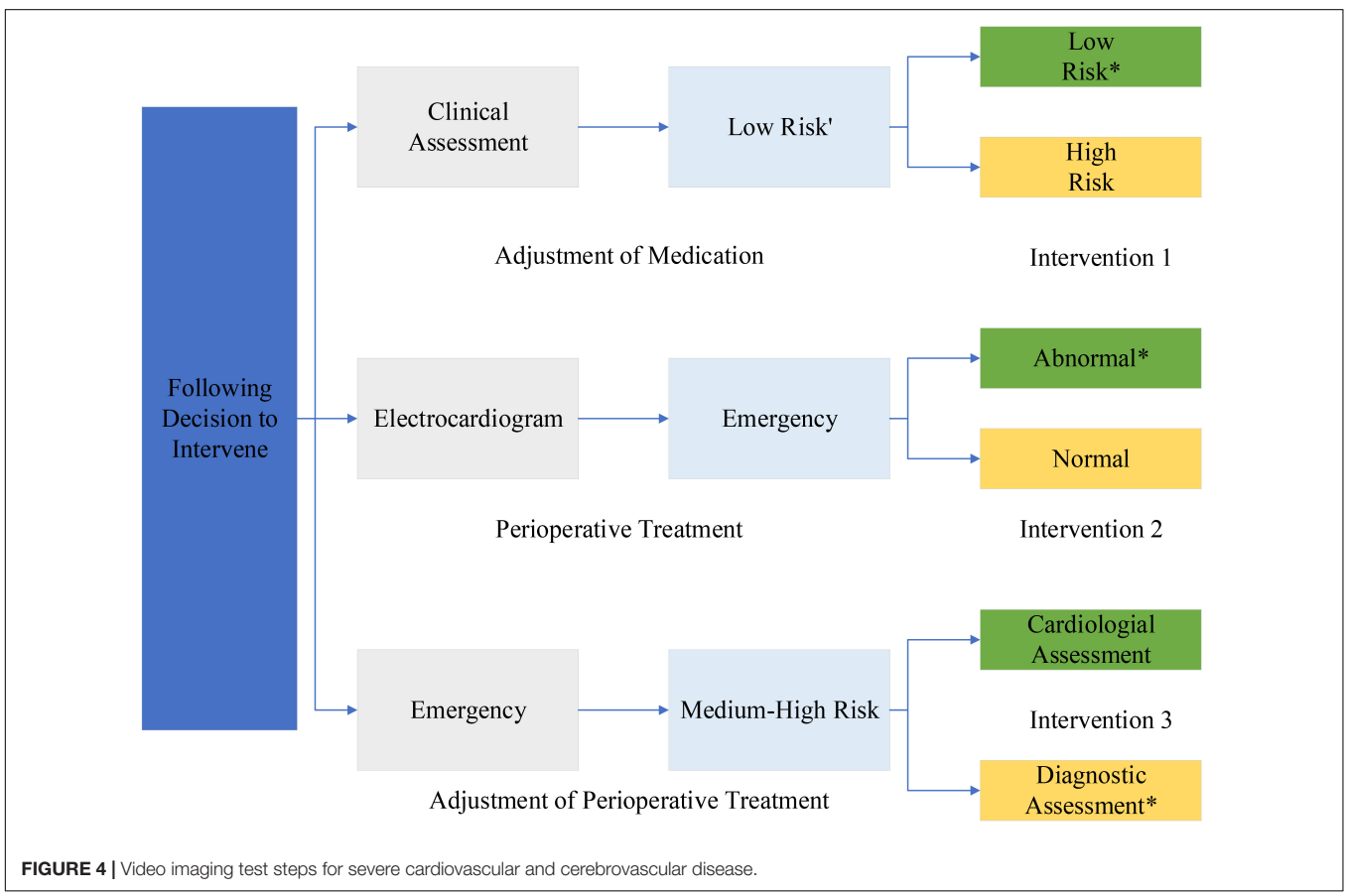


FIGURE 4 | Video imaging test steps for severe cardiovascular and cerebrovascular disease.

adapted to the pressure of the finger cuff. A small homemade cuff is wrapped around the middle phalanx of the finger, as shown in **Figure 4**. Continuous finger blood pressure is recorded for 15 cardiac cycles before the inflation of the cuff. The cuff is then rapidly inflated to 200 mmHg to close the finger vessels, and finger blood pressure signals are recorded for an additional 15 cardiac cycles as the pressure wave reaches a steady state after inflation. The ipsilateral brachial blood pressure then measured using a sociometric blood pressure device as a reference to upper arm blood pressure. An analogy band-pass filter with cut off frequencies of 1 and 35 Hz was also used to obtain the ECG signal. The ECG signal also obtained from the ECG module. Finally, a data acquisition system was used to simultaneously acquire the finger blood pressure and ECG signals at a sampling rate of 1,000 Hz.

The acquired data were processed offline, and if the mean arterial pressure difference (MAP) between the reference BAP and the continuous FAP was less than 4 mmHg, the BAP was reconstructed using Eqs (8) or (9) and the recorded data, where parameter A in Eq. (10) takes the mean value from the literature. The data processing is divided into three parts: pre-processing, calculation of PTT, and reconstruction of upper arm blood pressure, as shown in **Figure 4**. In the pre-processing section, the collected waveform is processed by 20 Hz, 5-step Barth low-pass filtering, and then the time markers are used to identify the valid data segments, i.e., 15 cycles before and after artery closure. In the PTT calculation section, the apex of the ECG R-wave and the start of the finger blood pressure wave before arterial closure are identified, then the Pulse Arrival Time (PAT) value of a single cycle is calculated and averaged over 15 cycles, then the Pre-Ejection Period (PEP) value is subtracted to obtain the total PTT from the finger to the heart. Finally, the PTT value from the finger to the upper arm is calculated proportionally. To reconstruct the upper arm blood pressure, we first adopt the method that the blood pressure wave measured at the time of artery closure is equal to two times of the advance wave to get the advance wave of blood pressure; then we subtract the advance wave from the finger blood pressure wave measured before artery closure to get the reflection wave and reflection coefficient of blood pressure; then we find the true reflection coefficient according to the relationship between the reflection coefficient before artery closure and the true reflection coefficient; then we can use the true blood pressure to obtain the reflection coefficient. The forward wave and reflection coefficient are used to derive the true blood pressure reflection wave, and finally, by adding the measured PTT parameters, we can reconstruct the true upper arm blood pressure waveform. In this section, two methods are used to reconstruct upper arm blood pressure: Method 1 uses the corrected reflection coefficient of the model in this section to reconstruct upper arm blood pressure, and Method 2 uses the uncorrected reflection coefficient to reconstruct upper arm blood pressure.

The model was divided into three levels, with the first level being the decision level, the second level being the guideline level, and the third level being the protocol level. The investigation carried out in a stepwise manner according to this model hierarchy from top to bottom. That is, the contents of the decision layer were first used as the basis for selecting the guideline layer evaluation score by making judgments by two-two comparisons, and then the contents of each guideline layer were used as the basis for selecting the protocol layer evaluation score, again by making judgments by two-two comparisons, respectively.

Collect clinical and imaging data including age, sex, and other risk factors for cerebrovascular disease and secondary prevention medications, and evaluate the patient's cranial magnetic resonance images including the specific location of the new infarct in the brainstem, whether it is located in the pars median penetrating artery supply area, whether it is located in multiple groups of brainstem penetrating artery supply areas, the size and volume of the new infarct, the SSS-TOAST classification, the severity of cerebral white matter loss, and the number of asymptomatic old luminal infarcts.

Analysis of Evaluation Indicators

The data extraction design to produce a uniform data extraction excel sheet required two researchers to independently read the literature, conduct an initial review and evaluation of the title and abstract of each randomized control study, and then read the full text for data extraction. Data extraction included: title of the literature, time of publication, experimental and control group information (total number, number of people, mean age, gender matching, jaded score, outcome measures). The review includes: inclusion and exclusion criteria of the literature, the setting of the control group, if there are disagreements can be discussed and resolved by the two researchers or assisted by other experts or researchers in the subject group, and the two researchers exchange checks after the work is completed. The methodological quality of the study is evaluated by two researchers alone for all included literature, generated from random sequences by applying the Jaded scale; concealment of random assignment scheme, specifically, both the trial implementer and the subjects before grouping The literature quality and methodological quality of the included literature was assessed in four areas: the specific allocation scheme of the subjects could not be known in advance; whether blinding was being used in the study; and whether withdrawal and withdrawal were being used in the study.

The video coding rate distortion algorithm, as a very important part of video coding, has become the subject of many researchers who have come up with many excellent ideas for improving the algorithm, with the ultimate goal of obtaining the best coding mode in which the bit rate R and coding distortion D make the coding cost J minimal. Under the H264 standard is each mode used for the current coding block. The resulting cost calculated for rate-distortion, and then comparative analysis of each mode is performed to pick the least costly class of coding modes, and the selected coding modes are defined as macroblocks to be coded. In H264, the RDO cost function shows that the substitution value is determined by three factors: motion search, reference frame selection, and mode decision, while the standard H264 only uses the traversal calculation method to optimize the rate-distortion, without fully considering the influence of other factors.

To solve this problem, this paper proposes its improvement ideas on how to confirm the inter-frame macroblock coding mode selection algorithm in the original rate-distortion

optimization. There are various ways to divide the inter-frame macroblock, according to the degree of motion we can make different divisions, the large division is suitable for absolute still or relatively still small amplitude motion, a small division is suitable for large changes in position and detail part of the more intense motion. The program finally adopts the 8*8 division of the block to be coded, divides the block into four 8*8 patterns, and analyses the corresponding motion vector direction for the four sub-modules obtained from the 8*8 pattern division. The set of possible coding patterns used in that coding block is then created together to create the coding pattern set and rate-distortion optimization is computed to get the best pattern needed, and tests show that the improved rate-distortion algorithm has improved the coding efficiency, as shown in **Figure 5**.

As the most accurate search algorithm, FS searches all the search points to get the best match, which is not suitable for transmission in the real-time video due to high computational complexity and can only be used as a standard for comparison with other algorithms. The three-step method is a simplification of the full search method, the fastest case only needs to search 25 search points to get the best match, although the complexity of the calculation is much reduced, and the matching accuracy is also reduced, the three-step method is only suitable for the application of a large range of motion of the frame image, and small-amplitude motion of the frame image, this algorithm is easy to fall into the state of the local optimal solution, which leads to large matching errors. The new three-step method is a supplement to the original three-step method, which makes use of the center bias feature to enhance the matching computation of the central region position, improve the search performance, and has good performance for video sequences with smaller motion. The hexagonal search method and the diamond search method, as the classical block-matching motion estimation algorithms, adopt the same idea and use two different search templates to avoid the defects of local optimization. The algorithm can be regarded as an improvement based on the hexagonal search method and the diamond search method, the hybrid search pattern with multiple templates combined with an early termination strategy can accurately predict the starting search point, which is recognized as the ideal motion estimation algorithm in H264.

ANALYSIS OF RESULTS

Analysis of Experimental Results

The matching matrix for the six classes of heartbeat classification is shown in **Figure 6**, where the rows represent the classification results of the six classes of heartbeats obtained by the algorithm, the columns represent the actual heartbeat classification results, and the data on the diagonal is the number of correctly detected heartbeats in each class of heartbeats. From the data in the table, there are more misidentifications between N beats and AP beats, N beats are easily recognized as AP beats and AP beats are easily recognized as N beats. The main reason is that N and AP beats are similar in morphology, the significant difference lies in the

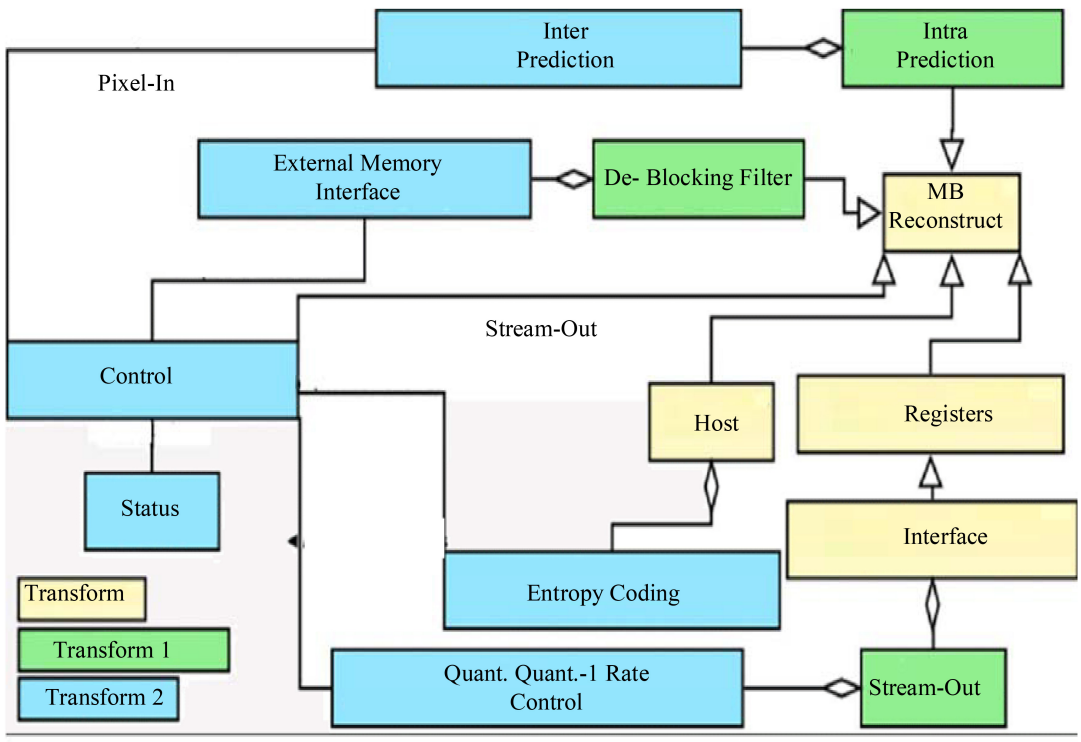


FIGURE 5 | Video image continuity evaluation.

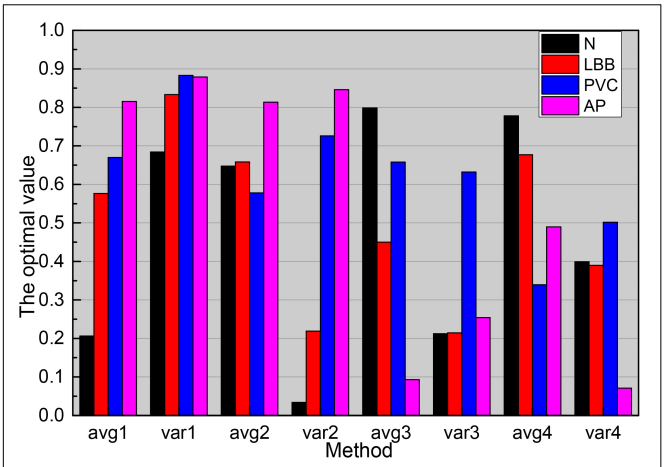


FIGURE 6 | Matching matrix for beat classification.

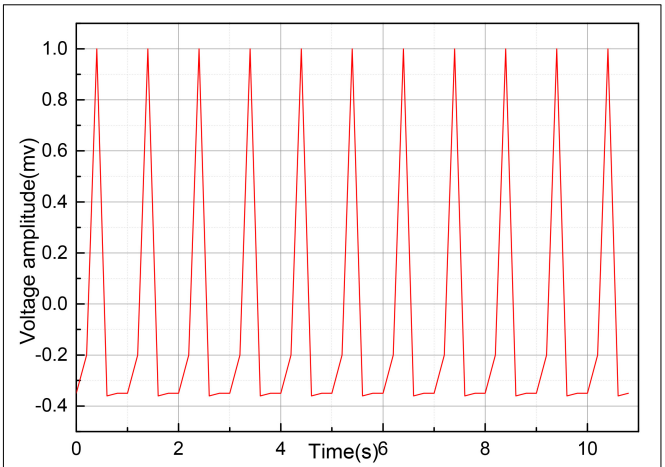


FIGURE 7 | Video image of ECG signal.

morphological changes of P-wave, which is a wave with a small amplitude, and the changes of P-wave amplitude are relatively large in different individuals, so it is easy to confuse N and AP beats. RBB beats also have some similarities with PVC beats in terms of morphology, which can also lead to misidentification, while P beats have good morphological consistency and are quite different from other types of beats, so P beats have the highest recognition rate.

Before the onset of SCD, the ECG signal pattern changes accordingly. **Figure 7** gives several groups of ECG signal morphology before the occurrence of SCD, the upper subgraph in each figure is the pre-processing signal, and the lower subgraph is the post-processing signal. As can be seen from the figure, although the signal morphology has some changes compared with the normal ECG signal, the morphological changes are not uniform, and the morphological changes are also different in different leads, and even the morphology is not consistent when SCD occurs in different people with the same lead data. In addition to the diversity of SCD evoked and the complexity of the noise contained in the ECG signal, accurate identification of SCD becomes difficult.

As the main parameters of the ESN network, the reserve pool connection radius, input scaling, and leakage rate have a great influence on the SCD identification results. It can be seen from the data in **Figure 8**, as the reserve pool connection spectrum radius decreases, its sensitivity, and prediction have increased, and the change is more obvious when it is reduced from 0.1 to 0.01, while when the reserve pool connection spectrum radius is further reduced, S and P no longer improve, instead, there is a significant decrease, so the final setting is 0.001. The input scaling controls the scaling of the input weight matrix. The scaling is too large to significantly reduce the recognition capability of the network, while too small will significantly reduce the recognition sensitivity, and finally 0.1. The leakage rate is the one in folium (4), which controls the rate of dynamic reserve pool update, and with the increasing value of m, the accuracy of SCD recognition is gradually improved until its value is close to 1.

The images were divided into five groups according to the characteristics of the algorithm: normal, border blur, stent, thrombus, and plaque groups. In the normal group, the images had an obvious layered structure with continuous highlighted areas outside the endothelial border; in the fuzzy border group, the images mainly contained those with an obscure layered structure, the similar grayscale value between layers, and extremely fuzzy border or border near the image edge; in the stent group, the images contained stent structure; in the thrombus group, the images contained obvious thrombus or residual blood in the intravascular lumen; and in the plaque group, the images contained various types of the plaque outside the intima-vascular membrane. In each group, 20 sets of images selected for analysis. **Figure 9** shows the endothelial extraction results of five groups of typical images, from top to bottom: normal, border blurred, stent, thrombus, and plaque. From left to right, the original images, the gold standard images, and the auto-extraction images of this algorithm are shown respectively. For five groups of typical images, this algorithm can accurately locate the endothelial border position and has a good agreement with the endothelium in the gold standard image.

To evaluate the effect of endomembrane extraction more accurately, we statistically analyzed the average sieve coefficient (D), Jaccard coefficient (J), Harsdorf distance (H), accuracy (P), and recall (R), and the corresponding mean squared deviation for each group of 20 images. At the same time, the change curves of D, J, P, and R for different images in the five sets of images are given, which are shown in **Figure 9**, where the red line is D, the blue line is J, the green line is P, and the black line is R. We can see the effect of the algorithm on the average sieve coefficient (D), Jaccard coefficient (J), Harsdorf distance (H), accuracy (P), and recall (R), and the corresponding mean squared deviation. We can see that the algorithm in this paper has the best effect on the endosomal extraction of normal images, and several evaluation indexes have good stability for different images. The main reason is that the normal group of

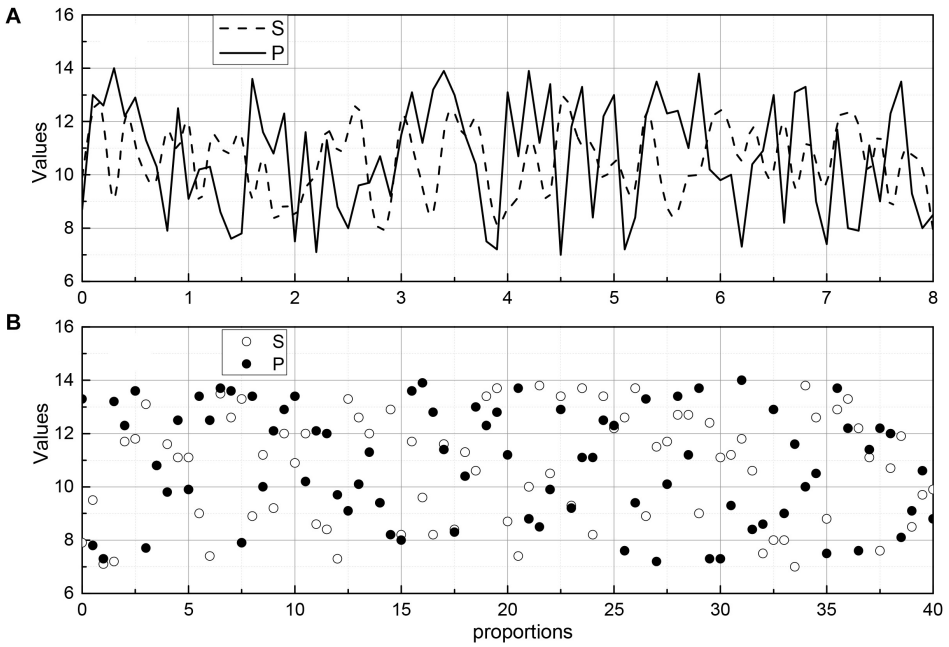


FIGURE 8 | Reserve pool connected spectral radius on the SCD recognition image.

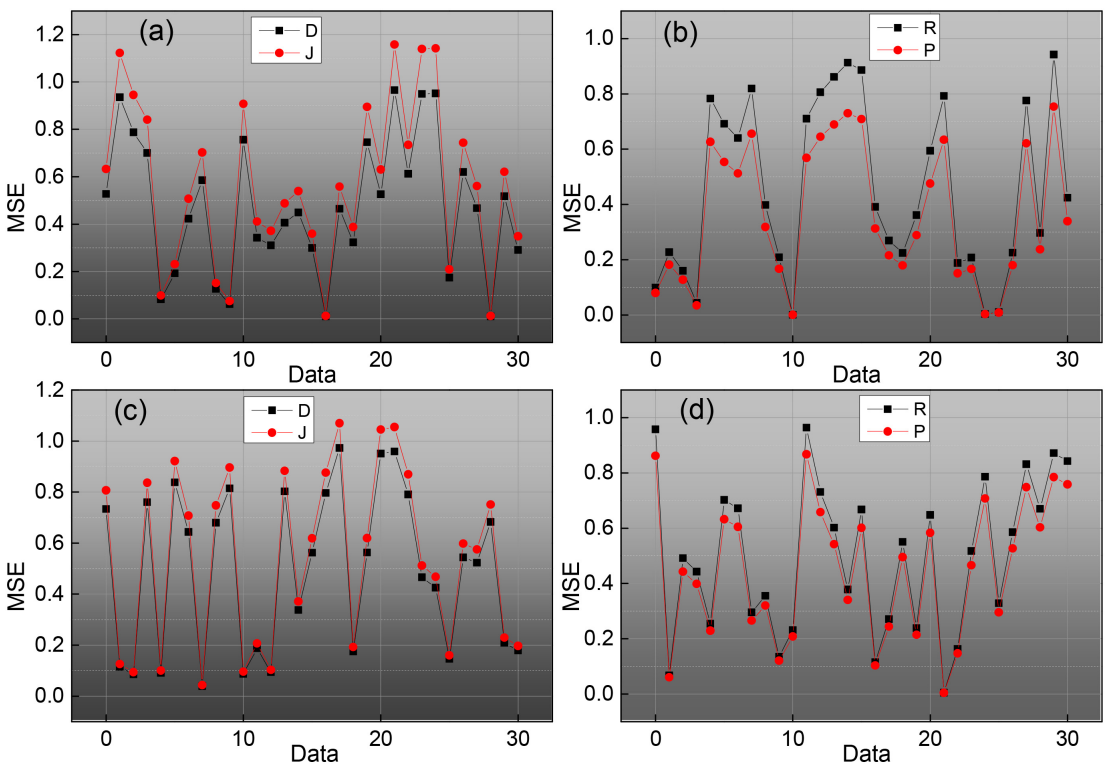


FIGURE 9 | Evaluation index value change curve.

images is relatively simple, and the images have high similarity, so it has high precision and stability. For other groups of images, due to the complexity of the image situation, the image itself is quite different, which makes the D, J, P, and R curves change a lot, and the average accuracy is also slightly lower than the normal group.

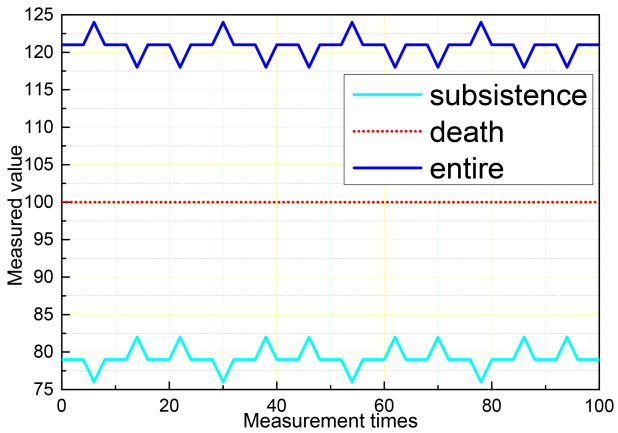


FIGURE 10 | Comparison of the three assessment methods with prognostic logistic regression analysis.

Artificial Intelligence Segmentation Dynamic Results Analysis

Logistic regression analysis was performed on aEEG classification and prognosis, and the χ^2-test was performed on the fitness of the regression equation model, $p = 0.003$, indicating that the model fit was good and the regression equation was significant. Logistic regression analysis showed that the accuracy of aEEG classification for survival prognosis was 100%, and the accuracy of the prognosis for death was 42.9%, and the accuracy of the judgment of the comprehensive prognosis was 88.6%. The results of logistic regression analysis showed that the accuracy of the EEG Young classification for survival prognosis was 96.4%, and the accuracy of the regression equation for death prognosis was 88.6%. The accuracy of the prognostic judgment was 57.1 and 88.6% for the combined prognostic judgment. The logistic regression analysis of the GCS score and prognosis showed that the accuracy of the GCS score for survival prognosis was 96.4% and the accuracy of the regression equation for death prognosis was 88.6%. 14.3% and the accuracy of the combined prognostic judgment was 80%, as shown in **Figure 10**. The results showed that aEEG score and EEG Young score were more accurate than the GCS score for prognostic judgment; and aEEG score combined with EEG Young score could improve the accuracy of prediction.

To illustrate more clearly the improvement of the effect brought by the method in this paper, this subsection gives a detailed data justification through experiments in terms of the training method, the parameters of the attention mechanism module, the number of fragments and sampling frames, and the effect brought by the fusion method, respectively, and makes a comparison with existing algorithms to analyze the advantages and disadvantages of the model. The training model of a neural network usually has three cases, one is to train from scratch, two is to use ImageNet pre-trained weights and fix the weights before the last layer so that they do not update the parameters, but only the parameters of the network behind, and three is to also use ImageNet pre-trained weights, but the weights are not fixed, but the parameters on the data set to be trained of fine-tuning. The three training methods are all based on the framework, and since the network does not have a pre-trained model, it is not pre-trained to train the spatial and temporal networks separately, and the experimental results are shown in **Figure 11**. As can be seen in **Figure 11**, the third training method

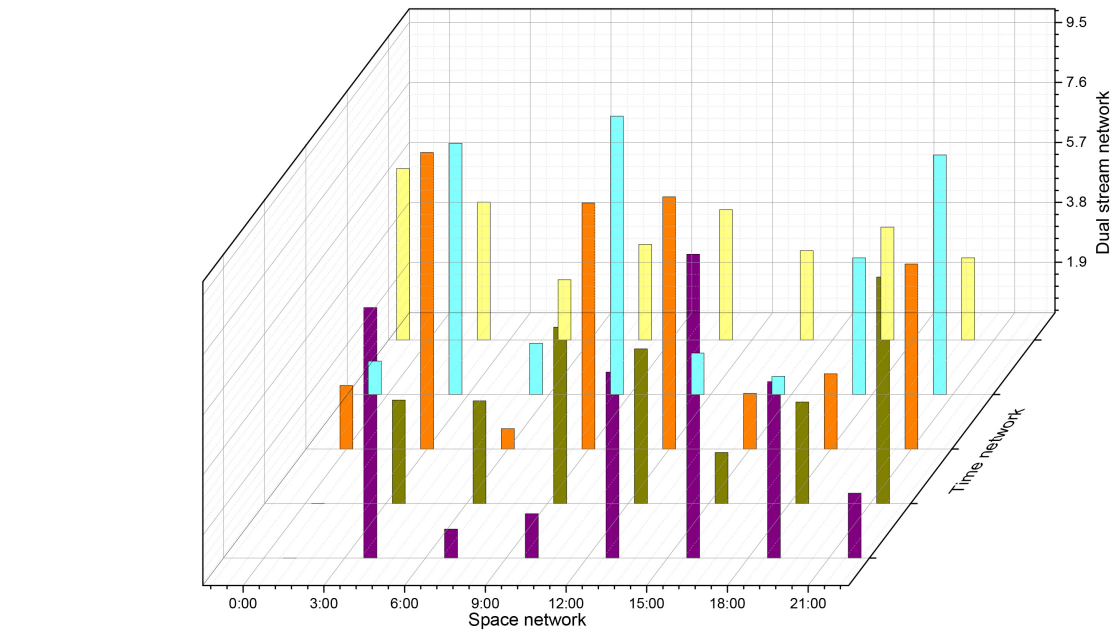


FIGURE 11 | Experimental accuracy on the dataset with different pre-training premises.

has the highest accuracy, training from zero has the worst effect, and fixing the previous parameters has the second-worst effect because training from zero is more difficult in the first place, it will take longer to converge when the gradient is updated, and it can converge to a satisfactory effect when the amount of data is sufficient, but the sample size of the dataset is small, only close to 10,000 training samples, so pre-training weights are needed, because ImageNet is a very large image dataset, on top of this training network weights for small datasets will speed up convergence, and improve accuracy, fixed pre-training weights, only update the network behind, but also compared to training from scratch has good results. So there are three ways to classify different scenarios and strengths, the first way to use the data set is relatively large, and no pre-training framework, this method of training convergence is slower, requiring a long time to see iterations to achieve better results for small data sets, this way the results are poor; the second way for small data sets, because the network parameters are large, small data sets cannot be on the network parameters are trained completely, in which case fixing the weights trained by ImageNet and updating only the last classification layer tends to achieve better results; the third way applies to cases such as data sets, which can be fine-tuned on the data set to be trained to achieve better accuracy and to speed up the convergence of the network. Given the above considerations, the third pre-training approach is chosen for all subsequent experiments.

The choice of the parameter time step of the attention mechanism module and its corresponding LSTM in the approach presented in this paper has a significant impact on the performance of the attention mechanism. The experiments in this section evaluate the accuracy of the action recognition obtained by the method at different TS values from 1 to 20, and the results on the databases are shown in **Figure 12**. When TS = 1, the distribution of attentional weights in 30 frames can be learned in only one iteration while fitting the distribution of importance in only 30 frames is not sufficient. As the TS value increases, the recognition accuracy gradually improves. When TS = 10, the proposed method achieves the highest recognition accuracy of 94.4 and 71.5% on the databases, respectively. As the value of TS continues to increase, the accuracy of action recognition decreases rapidly after remaining stable for some time. This is because as the TS value continues to increase, the size of the trainable parameters of the expanded LSTM increases rapidly, which leads to overfitting. Therefore, the time step TS of the attentional mechanism is determined to be 10 in this paper's approach.

To reduce the interference of catheter rings, residual blood in the vessel, or thrombus during A-line modeling, we use the endovascular boundary obtained by the algorithm in the previous chapter to expand the OCT image along the endovascular boundary. The specific implementation steps are to scan each column of the OCT image, detect the endothelial border points, and select the next point as the starting point of the column, and then take the points downward, a total of 100 pixels so that the OCT image is expanded along the endothelial border, as shown in **Figure 13**.

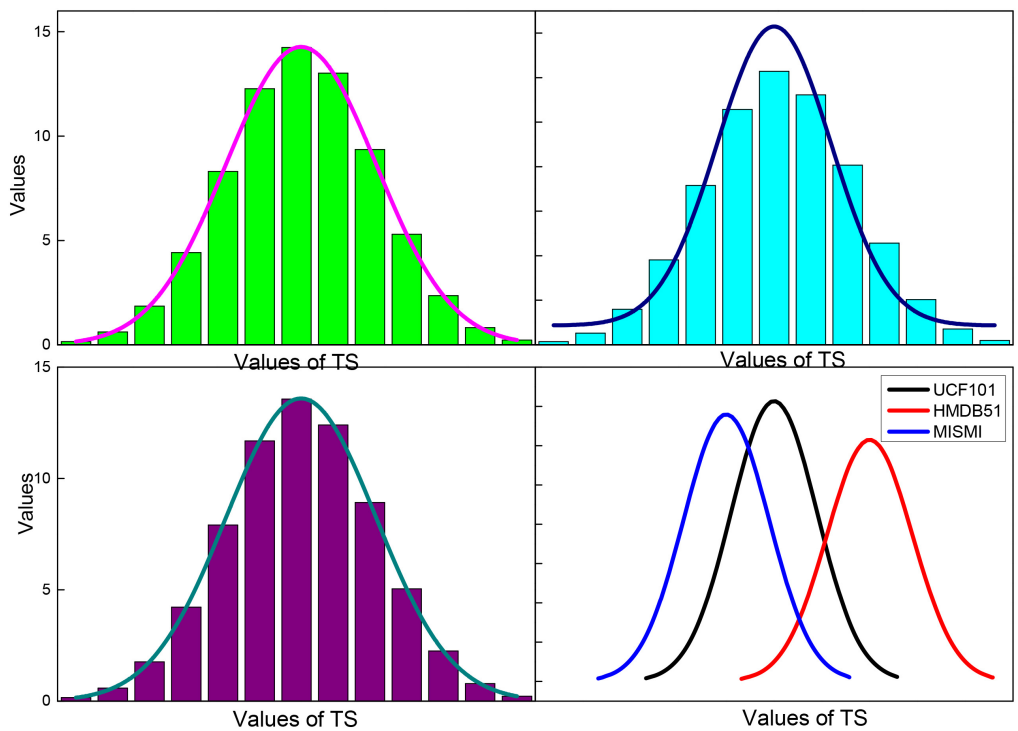


FIGURE 12 | Image recognition accuracy of the method in this paper for different TS values.

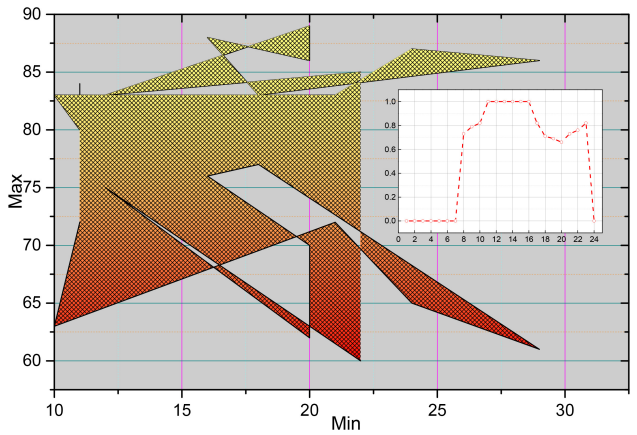


FIGURE 13 | Unfolded image and extract A-line.

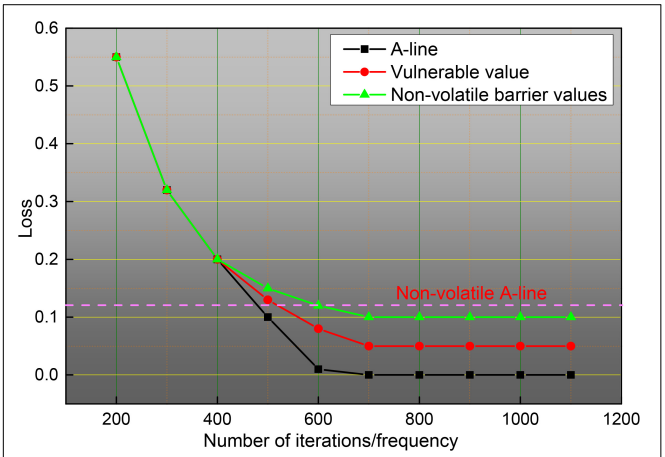


FIGURE 14 | Perishable and non-perishable A-line.

In the unfolded OCT image, each column element constitutes an A-line, which has different morphological characteristics of F due to the light absorption property of the plaque itself and the position relation with the intima. **Figure 14** shows four typical A-line morphologies, in which the red line is F, the green line is F-C, the blue line is F-L, and the pink line is GS. Due to the clear border of calcified plaques, the A-line of F-C plaques will have a more drastic change in slope; due to the blurred border of F-L plaques, the A-line of F-L plaques will usually be relatively flat, and there will be a staggered distribution of high and low values; while the A-line of GS plaques is close to the zero line, and there may be small fluctuations at the starting position.

In this paper, we focus on intelligent analysis methods for high-risk cardiovascular diseases, making full use of the portable and long acquisition time features of dynamic ECG to study the intelligent prediction algorithm for arrhythmias and SCD at the level of ECG signal. A sparse auto-coding deep neural network with a four layer stack structure was constructed to automatically extract the depth features of arrhythmia beats, the automatic recognition of six classes of beats was achieved by SoftMax classifier, the training process of the network was optimized by a parsimonious Newtonian optimization algorithm so that the extracted depth features could more accurately describe the input signal, and the key parameters of the network were discussed and analyzed. By taking full advantage of the recognition ability of the echo state network on the time domain signal, the ESN network with a multilayer tandem structure was designed to achieve accurate recognition of the sudden cardiac death signal. The accurate predictions were 93.04, 95.36, 94.20, 94.20, and 94.78%, respectively, when tested using the signal 5 min before the onset of sudden death.

CONCLUSION

In this paper, the continuity of segmented dynamic video images in the detection of severe cardiovascular and cerebrovascular diseases is investigated and analyzed by artificial intelligence, taking full advantage of the high precision and high resolution of medical images. Based on the characteristics of different OCT images containing different tissue information with different gray level distribution, and automatic gray level label value selection method based on image gray level distribution characteristics developed to ensure the stability of the algorithm in the extraction of endosomes from different OCT images. The effect of the algorithm on the endomembrane extraction under different disturbances is also verified. Fully utilizing the self-learning ability of stack sparse automatic coding network for label-free data, the deep modeling analysis of A-line in OCT images was performed to achieve accurate identification of fibro genic, fibro genic-calcified, and fibro genic-lipid A-line, and the automatic extraction of plaque regions based on the automatic plaque region generation algorithm. On this basis, the automatic identification of thin fibrous cap-like vulnerable plaques was achieved by analyzing the fibrous cap thickness. In this paper, the algorithm achieved simultaneous automatic recognition of plaque and vulnerable plaque by modeling and analysis of A-line in OCT images, with recognition accuracy, recall rate, and area overlap of 84.29, 84.29, and 87.33%, respectively.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html.

AUTHOR CONTRIBUTIONS

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

REFERENCES

Ali, S., Ismail, Y., and Badawi, A. (2018). Adaptive multi-connection DASH scalable video coding for wireless area networks. *Int. J. Comput. Digital Syst.* 7, 175–188. doi: 10.12785/ijcds/070401

Aminikhanghahi, S., and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowl. Inform. Syst.* 51, 339–367. doi: 10.1007/s10115-016-0987-z

Banerjee, D., Islam, K., Xue, K., Mei, G., Xiao, L., Zhang, G., et al. (2019). A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. *Knowl. Inform. Syst.* 60, 1693–1724. doi: 10.1007/s10115-019-01337-2

Bechar, M. E. A., Settouti, N., Barra, V., and Chikh, M. A. (2018). Semi-supervised superpixel classification for medical images segmentation: application to detection of glaucoma disease. *Multidimensional Syst. Signal Process.* 29, 979–998. doi: 10.1007/s11045-017-0483-y

Costa, A. C., Ahamed, T., and Stephens, G. J. (2019). Adaptive, locally linear models of complex dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 116, 1501–1510. doi: 10.1073/pnas.1813476116

Craye, C., Rashwan, A., Kamel, M. S., and Karray, F. (2016). A multi-modal driver fatigue and distraction assessment system. *Int. J. Intellig. Transport. Syst. Res.* 14, 173–194. doi: 10.1007/s13177-015-0112-9

Darwish, A., Hassanien, A. E., and Das, S. (2020). A survey of swarm and evolutionary computing approaches for deep learning. *Artif. Intellig. Rev.* 53, 1767–1812. doi: 10.1007/s10462-019-09719-2

Deldari, S., Smith, D. V., Sadri, A., and Salim, F. (2020). ESPRESSO: entropy and ShaPe awaRe timE-Series SegmentatiOn for processing heterogeneous sensor data. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 4, 1–24. doi: 10.1145/3411832

Diehn, F. E., Maus, T. P., Morris, J. M., Carr, C. M., Kotsenas, A. L., Luetmer, P. H., et al. (2016). Uncommon manifestations of intervertebral disk pathologic conditions. *Radiographics* 36, 801–823. doi: 10.1148/rg.2016150223

Eldib, M., Deboeverie, F., Philips, W., and Aghajan, H. (2018). Discovering activity patterns in office environment using a network of low-resolution visual sensors. *J. Ambient Intellig. Hum. Comput.* 9, 381–411. doi: 10.1007/s12652-017-0511-7

Eriksson, P. E., and Eriksson, Y. (2019). Live-action communication design: a technical how-to video case study. *Techn. Commun. Q.* 28, 69–91. doi: 10.1080/10572252.2018.1528388

Fan, C. L., Lo, W. C., Pai, Y. T., and Hsu, C. H. (2019). A survey on 360 video streaming: acquisition, transmission, and display. *ACM Comput. Surv.* 52, 1–36. doi: 10.1145/3329119

Haider, F., De La Fuente, S., and Luz, S. (2019). An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech. *IEEE J. Select. Top. Signal Process.* 14, 272–281. doi: 10.1109/jstsp.2019.2955022

Hong, L., Luo, M., Wang, R., Lu, P., Lu, W., and Lu, L. (2018). Big data in health care: applications and challenges. *Data Inform. Manag.* 2, 175–197. doi: 10.2478/dim-2018-0014

Horton, M. B., Brady, C. J., Cavallerano, J., Abramoff, M., Barker, G., Chiang, M. F., et al. (2020). Practice guidelines for ocular telehealth-diabetic retinopathy. *Telemed. e-Health* 26, 495–543. doi: 10.1089/tmj.2020.0006

Humphreys, A., and Wang, R. J. H. (2018). Automated text analysis for consumer research. *J. Consum. Res.* 44, 1274–1306. doi: 10.1093/jcr/ucx104

Jeannot, G. (2019). Smart city projects in the continuity of the urban socio-technical regime: the french case. *Inform. Polity* 24, 325–343. doi: 10.3233/ip-190128

Latha, S., Samiappan, D., and Kumar, R. (2020). Carotid artery ultrasound image analysis: a review of the literature. *Proc. Instit. Mech. Eng. H J. Eng. Med.* 234, 417–443.

Lee, J. H., Yoon, S. H., Kim, Y. T., Kang, C. H., Park, I. K., Park, S., et al. (2019). Sleeve lobectomy for non–small cell lung cancers: predictive CT features for resectability and outcome analysis. *AJR* 213, 807–816. doi: 10.2214/ajr.19.21258

Luo, Y., Yang, B., Xu, L., Hao, L., Liu, J., Yao, Y., et al. (2018). Segmentation of the left ventricle in cardiac MRI using a hierarchical extreme learning machine model. *Int. J. Mach. Learn. Cybernet.* 9, 1741–1751. doi: 10.1007/s13042-017-0678-4

Mariakakis, A., Baudin, J., Whitmire, E., Mehta, V., Banks, M. A., Law, A., et al. (2017). PupilScreen: using smartphones to assess traumatic brain injury. *Proc. ACM Interact., Mobile Wearable Ubiquitous Technol.* 1, 1–27. doi: 10.1145/3131896

Musial, G., Queener, H. M., Adhikari, S., Mirhajianmoghadam, H., Schill, A. W., Patel, N. B., et al. (2020). Automatic segmentation of retinal capillaries in adaptive optics scanning laser ophthalmoscope perfusion images using a convolutional neural network. *Transl. Vis. Sci. Technol.* 9:43. doi: 10.1167/tvst.9.2.43

Nascimento, J. C., and Carneiro, G. (2017). Deep learning on sparse manifolds for faster object segmentation. *IEEE Trans. Image Process.* 26, 4978–4990. doi: 10.1109/tip.2017.2725582

Pasterkamp, H. (2018). The highs and lows of wheezing: a review of the most popular adventitious lung sound. *Pediatr. Pulmonol.* 53, 243–254. doi: 10.1002/ppul.23930

Randive, S. N., Senapati, R. K., and Rahulkar, A. D. (2020). Intelligent model for diabetic retinopathy diagnosis: a hybridised approach. *Int. J. Bioinform. Res. Appl.* 16, 120–150. doi: 10.1504/ijbra.2020.10030363

Ren, S., Sun, K., Liu, D., and Dong, F. (2019). A statistical shape-constrained reconstruction framework for electrical impedance tomography. *IEEE Trans. Med. Imaging* 38, 2400–2410. doi: 10.1109/tmi.2019.2900031

Saggi, M. K., and Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Inform. Process. Manag.* 54, 758–790. doi: 10.1016/j.ipm.2018.01.010

Salekin, A., Eberle, J. W., Glenn, J. J., Teachman, B. A., and Stankovic, J. A. (2018). A weakly supervised learning framework for detecting social anxiety and depression. *Proc. ACM interact. Mobile Wearable Ubiquitous Technol.* 2, 1–26. doi: 10.1145/3214284

Shinbane, J. S., and Saxon, L. A. (2018). Virtual medicine: utilization of the advanced cardiac imaging patient avatar for procedural planning and facilitation. *J. Cardiovasc. Comput. Tomogr.* 12, 16–27. doi: 10.1016/j.jcct.2017.11.004

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic stimuli in neuroscience: critically acclaimed. *Trends Cogn. Sci.* 23, 699–714. doi: 10.1016/j.tics.2019.05.004

Xiao, D., Bhuiyan, A., Frost, S., Vignarajan, J., Tay-Kearney, M. L., and Kanagasingam, Y. (2019). Major automatic diabetic retinopathy screening systems and related core algorithms: a review. *Mach. Vis. Appl.* 30, 423–446. doi: 10.1007/s00138-018-00998-3

Yang, S., Li, B., Cao, Y. P., Fu, H., Lai, Y. K., Kobbelt, L., et al. (2020). Noise-resilient reconstruction of panoramas and 3D scenes using robot-mounted unsynchronized commodity RGB-D cameras. *IEEE Trans. Visualization Comput. Graph.* 26, 2485–2498.

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 15 March 2021
doi: 10.3389/fnins.2021.655019

Research on the Modality Transfer Method of Brain Imaging Based on Generative Adversarial Network

Dapeng Cheng[1,2]*, *Nuan Qiu*[1], *Feng Zhao*[1,2], *Yanyan Mao*[1,2] and *Chengnuo Li*[1]

[1] School of Computer Science and Technology, Shandong Technology and Business University, Yantai, China, [2] Shandong Co-Innovation Center of Future Intelligent Computing, Yantai, China

Brain imaging technology is an important means to study brain diseases. The commonly used brain imaging technologies are fMRI and EEG. Clinical practice has shown that although fMRI is superior to EEG in observing the anatomical details of some diseases that are difficult to diagnose, its costs are prohibitive. In particular, more and more patients who use metal implants cannot use this technology. In contrast, EEG technology is easier to implement. Therefore, to break through the limitations of fMRI technology, we propose a brain imaging modality transfer framework, namely BMT-GAN, based on a generative adversarial network. The framework introduces a new non-adversarial loss to reduce the perception and style difference between input and output images. It also realizes the conversion from EEG modality data to fMRI modality data and provides comprehensive reference information of EEG and fMRI for radiologists. Finally, a qualitative and quantitative comparison with the existing GAN-based brain imaging modality transfer approaches demonstrates the superiority of our framework.

Keywords: brain imaging modality transfer, EEG, fMRI, generative adversarial network, non-adversarial loss

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Lin Wang,
Yanshan University, China
Rui Li,
Xidian University, China

***Correspondence:**
Dapeng Cheng
chengdapeng@sdtbu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *18 January 2021*
Accepted: *22 February 2021*
Published: *15 March 2021*

Citation:
Cheng D, Qiu N, Zhao F, Mao Y and Li C (2021) Research on the Modality Transfer Method of Brain Imaging Based on Generative Adversarial Network. Front. Neurosci. 15:655019. doi: 10.3389/fnins.2021.655019

1. INTRODUCTION

Brain imaging modality transfer has become popular in the field of medical imaging, providing a variety of reference information for early diagnosis, identification, treatment, and follow-up of the diseases (Yi et al., 2019). For example, the translation of computed tomography (CT) modality data to magnetic resonance (MR) modality data (Jin et al., 2019) solves the problem of lack of MR modality data or synthesizes positron emission tomography (PET) modality data from CT modality data (Choi and Lee, 2018) for cancer staging, detection, and treatment.

The modality transfer of brain imaging is to map one modality to another (Armanious et al., 2020). The information provided by an image obtained from a certain imaging method is often limited, and can only reflect the information of one modality, which generally cannot help the doctor to make an accurate diagnosis. The modality transfer technology is conducive to converting between different modality images to obtain multi-modality information. Combined with multi-modality images, it can provide a variety of information about diseased tissues or organs, and provide a strong theoretical basis for clinical medicine to make an accurate diagnosis. As for the modality transfer of brain imaging, the current research at home and abroad is mainly based on the image-to-image translation. Existing image-to-image translation methods are mainly divided into two categories: sparse representation-based method and learning-based method.

The method based on sparse representation (SR) is the process of using image block prediction. An image block is extracted from an atlas with the same modalities as the source image to sparsely

represent the source image, and then the obtained sparse coefficient is used to estimate the target image block of another modality image. Ye et al. (2013) used T1-weighted MR images to synthesize T2-weighted MR images. Cordier et al. (2016) used brain labeled images to predict FLAIR images. However, the sparse representation-based approach requires sparse coding optimization of image blocks at all locations, which will lead to a decrease in the efficiency of the prediction process, and sometimes there will be noise in the synthesis results.

The convolutional neural network (CNN) in the "two networks" shows good performance in the fields of computer vision and medical image analysis. Li et al. (2014) used CNN to predict PET images from MR. Han (2017) used deep CNN to synthesize CT with MR. But, the training of a robust CNN model requires a large data set and a long training time. However, there is an unavoidable problem in the CNN-based image modality migration task. Although the entire network training process is automatic, it still needs to design loss function, namely, design loss network parameters.

Since the GAN concept was proposed, it has been showing strong practicability in image synthesis. The work of Denton et al. (2015) and Salimans et al. (2016) has proved that GAN has been showing strong practicability in the image. GAN can solve the problems encountered in CNN well. According to the work of Zhao et al. (2016), it is proved that when using CNN, a specific loss function is needed to deal with the problem, and GAN can achieve good results. At the same time, a new GAN network structure—Conditional GANs was proposed, referring to the design pattern of the CNN network and adding Settings for specific applications when GAN was used. A lot of work confirmed that such a GAN network has a good effect in processing images, videos, and 3D data. Isola et al. (2017) have proposed an image translation structure Pix2Pix based on conditional GAN, which can achieve good results under supervised learning.

However, a network based on supervised learning requires paired data training, and patients should be tested for both EEG and fMRI at short intervals. Besides, further, post-processing is required, such as image registration (Yang et al., 2018). Obtaining these data is a great challenge. Whatsmore, with the development of computer technology and digital imaging technology, brain imaging methods are gradually mature and diversified, yet many of these technologies still have limitations. For example, functional magnetic resonance imaging types of equipment are expensive, and patients who use pacemakers cannot use the fMRI. Although fMRI can accurately locate the active area when "observing the active brain" and its spatial resolution can reach the millimeter level (Tang et al., 2019), its temporal resolution is far lower than EEG (Menon et al., 1998). In contrast, EEG is easier to implement. Thus, if we can realize the conversion from EEG to fMRI, we can obtain both advantages of EEG and fMRI by implementing only one of them. However, to our best knowledge, few studies have explored the conversion of EEG to fMRI modality data.

In this work, we propose a brain imaging modality transfer framework, which uses 2D sensor cap images of EEG and T1-weighted axial images from fMRI to infer the target fMRI modality data, providing comprehensive reference information for medical diagnosis. Specifically, we focus on generating fMRI modality data from EEG modality data, which requires learning the mapping relationship between these two different modality data. However, it is well-known that converting between different modality data has always been a challenge. Fortunately, the emergence of the generative adversarial network (GAN) (Goodfellow et al., 2014) improves this aspect. GAN has achieved advanced performance in learning the mapping between different modalities. Meanwhile, with the development of deep learning, GAN has become one of the hot research directions in the field of medical imaging.

Therefore, we leverage GAN to achieve a brain imaging modality transfer framework named BMT-GAN. This framework introduces the combination of cycle-consistency loss and adversarial loss used in the CycleGAN framework proposed by Wolterink et al. (2017). which avoids the inability of traditional GAN to achieve pairing problems between input and output. Besides, the framework introduces non-adversarial loss, which reduces the perception and style differences between input and output modality images, and enhances the global consistency between input and output images. A qualitative and quantitative comparison with the existing GAN-based medical image modality transfer methods proves that BMT-GAN can effectively capture the overall tissue contrast and local anatomical details of fMRI images.

The contributions of this article are summarized as follows:

(1) We proposed a new framework named BMT-GAN to estimate two-dimensional fMRI images using two-dimensional EEG images;

(2) BMT-GAN framework combines the cycle-consistent loss, adversarial loss, and non-adversarial loss to achieve excellent brain imaging modality transfer performance;

(3) The proposed approach can be easily extended to other medical data translation tasks to benefit the medical imaging field.

2. METHODS

The proposed BMT-GAN framework utilizes the dual thought (Farnia and Tse, 2018) of Welander et al. (2018), which not only adopts the cycle-consistency loss and adversarial loss of CycleGAN model but also introduces a new non-adversarial loss combination. Whatsmore, **Figures 1A,B** shows the BMT-GAN framework structure of two parts: CycleGAN and non-adversarial structure.

2.1. CycleGAN

CycleGAN is an unsupervised framework that allows image-to-image translation without matching training data samples in two domains. The network framework structure and data flow diagram are shown in **Figure 1A**.

As shown in **Figure 1A**, the differences between CycleGAN framework and traditional GAN are: (1) CycleGAN consists of two generators (G_1, G_2) and two discriminators (D_1, D_2); (2) CycleGAN instead of learning a single mapping, it learns two mapping functions between two domains: $G_1 : I_{EEG} \longrightarrow$

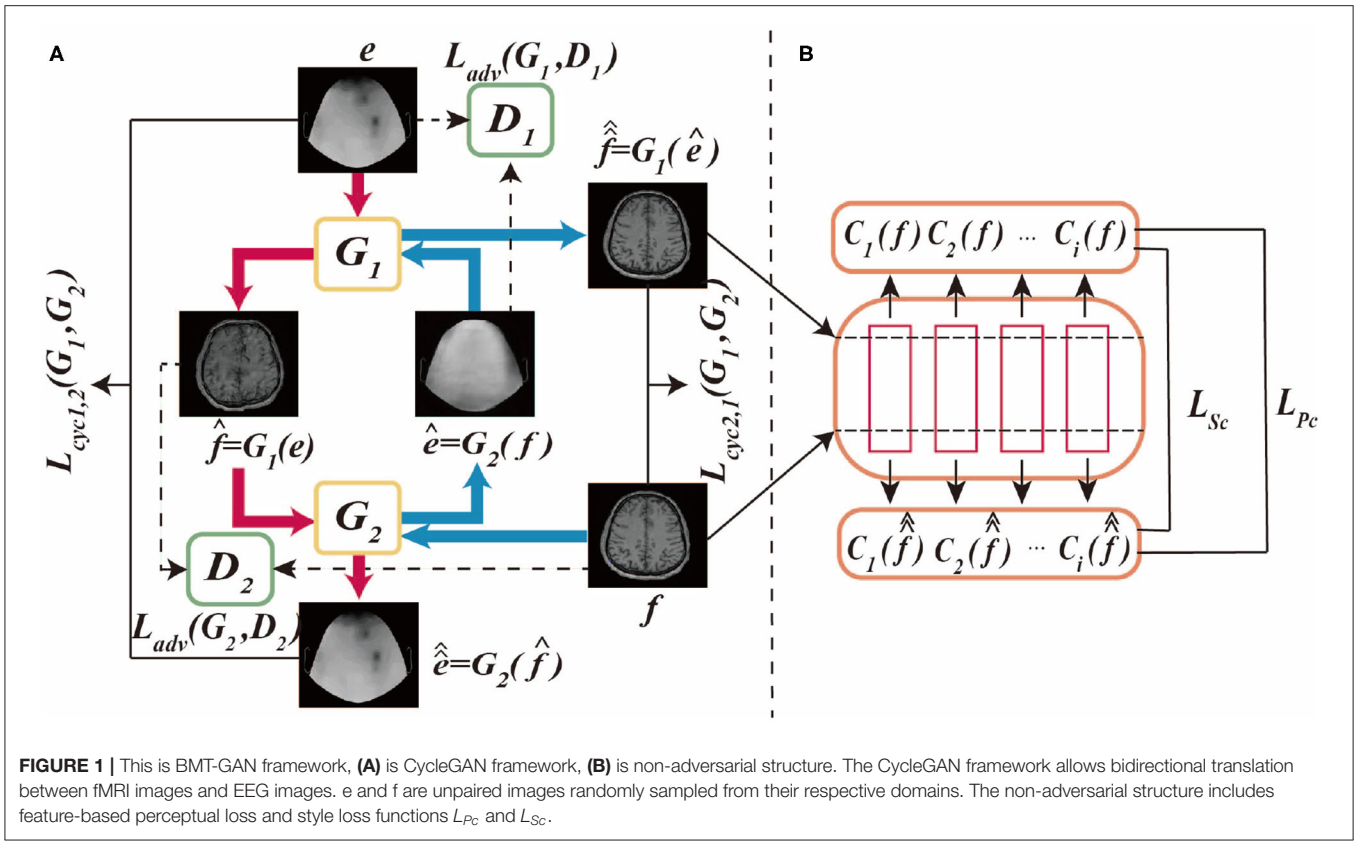


FIGURE 1 | This is BMT-GAN framework, **(A)** is CycleGAN framework, **(B)** is non-adversarial structure. The CycleGAN framework allows bidirectional translation between fMRI images and EEG images. e and f are unpaired images randomly sampled from their respective domains. The non-adversarial structure includes feature-based perceptual loss and style loss functions L_{Pc} and L_{Sc}.

I_{fMRI} and $G_2 : I_{fMRI} \longrightarrow I_{EEG}$, ensuring that the input and output images can be matched after translation. Where, I_{EEG} is the sample space of the given EEG image training data, $e \in I_{EEG}$, I_{fMRI} is the sample space of the given fMRI image training data, $f \in I_{fMRI}$.

The framework of CycleGAN consists of two major branches $Cyc_{1,2}$ and $Cyc_{2,1}$. In the generation network of the $Cyc_{1,2}$ branch, the generator G_1 takes the EEG images in the source domain I_{EEG} as input, and outputs the synthetic translation $\hat{f} = G_1(e)$. To prevent a group of different images from being mapped to a single image in the target domain, the model requires the two generators to keep the loop consistent with each other. Therefore, using $\hat{f}$ as the input of G_2, the output reconstruction $\hat{\hat{e}} = G_2(\hat{f})$ is obtained. In the $Cyc_{2,1}$ branch, the discriminator uses the fMRI images of the target domain I_{fMRI} as the input of G_2, and the synthesized translation $\hat{e} = G_2(f)$. In the same way, using $\hat{e}$ as the input of G_1 to get the output reconstruction $\hat{\hat{f}} = G_1(\hat{e})$. Therefore, the cycle-consistent loss of the two branches can be formulated as:

$$\begin{aligned} L_{Cyc_{1,2}}(G_1, G_2) &= E_{e\in I_{EEG}}[||G_2(G_1(e)) - e||_1] \\ L_{Cyc_{2,1}}(G_2, G_1) &= E_{f\in I_{fMRI}}[||G_1(G_2(f)) - f||_1] \end{aligned} \tag{1}$$

Also, each generator uses a corresponding discriminator for adversarial training. In the discriminator of the $Cyc_{1,2}$ branch, the fMRI image in the target domain I_{fMRI} and the synthetic translation $\hat{f}$ are used as the D_1 input of the discriminator. In the $Cyc_{2,1}$ branch, the EEG image in the source domain I_{EEG} and the synthetic translation $\hat{e}$ from the source domain are used as the input of D_2. Among them, D_1 and D_2 act as binary classifiers to distinguish the converted image from the target domain image as accurately as possible. Besides, the generator synthesizes high-quality images through iterative training to obfuscate the discriminator. Therefore, the model's adversarial loss function can be formulated as:

$$\begin{aligned} L_{Adv_{1,2}}(G_1, D_1) &= E_{f\in I_{fMRI}}[logD_1(f)] \\ &\quad + E_{e\in I_{EEG}}[log(1 - D_1(G_1(e)))] \\ L_{Adv_{2,1}}(G_2, D_2) &= E_{e\in I_{EEG}}[logD_2(e)] \\ &\quad + E_{f\in I_{fMRI}}[log(1 - D_2(G_2(f)))] \end{aligned} \tag{2}$$

2.2. Non-adversarial Structure

In order to ensure that the input and output images can be paired after image translation, the loss functions in Equations (1) and (2) are introduced. However, during the training process, due to the large differences in the anatomical details of the 2D EEG image and the 2D fMRI image, the pixel-level loss becomes ineffective and the output translation image lacks clarity and fine structure (Johnson et al., 2016; Wang C. et al., 2018). For example, Jin et al. (2019) used a combination of adversarial loss and the traditional L1 distance loss. This loss combination has achieved good results in the translation task of paired images. Armanious et al. (2020)

used a combination of perceptual loss and style loss in addition to cycle-consistent and adversarial loss in the MedGAN framework, and also achieved success.

However, for unpaired image translation tasks, it is not feasible to use the above two-loss combinations due to the huge difference between the source domain image and the target domain image. Therefore, we introduce a feature-based loss function between input reference and output reconstruction as an additional constraint to improve output quality. The process is shown in **Figure 1B**. The first proposed loss function is the cycle-perceptual loss, L_{Pc}. This is achieved by extracting intermediate feature maps using a pre-trained feature extractor network, for both the input and the cycle-reconstructed images. This process is formulated as a non-adversarial perceptual loss function L_{Pc}:

$$L_{Pc} = \sum_{i=0}^{N} \omega_{pc,i}(||C_i(e) - C_i(\hat{\hat{e}})||_1 + ||C_i(f) - C_i(\hat{\hat{f}})||_1) \quad (3)$$

Where C_i is the feature map extracted by the *ith* layer of the feature extractor network. N is the total number of layers, and $\omega_{pc,i}$ is the weight of each layer.

Secondly, a style-based loss function between input reference and output reconstruction is proposed to compensate for the difference in style representation between the reconstructed image and its corresponding target image. By calculating the relevance of the feature representation in the spatial range, the style distribution can be captured. To calculate the style loss,we must first calculate $G_{r_i}(e)$:

$$G_{r_i}(e)_{m,n} = \frac{1}{h_i, w_i, d_i} \sum_{h=1}^{h_i} \sum_{w=1}^{w_i} C_i(e)_{h,w,m} C_i(e)_{h,w,m} \quad (4)$$

Where $G_{r_i}(e)$ is the Gram matrix of each convolution block, the shape is $d_i\ d_i$, which is used to express the feature association. Its elements are obtained by inner product calculation on the height and width of the feature graph. h_i, w_i, d_i the height, width, and depth of the feature space. The non-adversarial style loss function is formulated as L_{Sc}:

$$L_{Sc} = \sum_{i=0}^{N} \omega_{sc,i} \frac{1}{4d_i^2} (||G_{r_i}(e) - G_{r_i}(\hat{\hat{e}})||_F^2 + ||G_{r_i}(f) - G_{r_i}(\hat{\hat{f}})||_F^2) \quad (5)$$

$\omega_{sc,i}$ is the ith weight of the given Gram matrix.

Therefore, the ultimate minimum and maximum optimization task of the BMT-GAN framework is formulated as follows:

$$\min_{G_1,G_2} \max_{D_1,D_2} L = L_{Cyc_{1,2}}(G_1, G_2) + L_{Cyc_{2,1}}(G_2, G_1) + L_{Adv_{1,2}}(G_1, D_1) + L_{Adv_{2,1}}(G_2, D_2) + \omega_{pc} L_{Pc} + \omega_{sc} L_{Sc} \quad (6)$$

2.3. Generator and Discriminator

The generator network is composed of three components: encoder, converter, and decoder, while the discriminator network is composed of multiple convolutional layers, the flow diagram as shown in **Figures 2A,B**.

2.3.1. Generator Network

(1) The encoder is composed of three convolutional layers. The initial size of the input image is 256 × 256 × 1. The first step is to extract features from the image through the convolutional layer. The number of features extracted from the convolutional layer can also be regarded as the number of different filters used to extract different features. The convolutional layer gradually extracts more advanced features in turn. After passing through the encoder, the input image changes from (256,256,1) to (64,64,128).

(2) The converter is composed of nine residual blocks. Different features of the image are combined by different channels of the output image of the encoder. According to these features, the feature vector of the image is converted from the source domain to the target domain. However, since the images in the EEG domain and fMRI domain do not have similar features, the feature vectors of images in the EEG domain should be converted into feature vectors in the fMRI domain through the sharing function of the generator during the conversion process. The residual block consists of two convolutional layers and the input residuals are added to the output to ensure that the input properties of the previous layer can also be applied to the later layers so that their output will not be different from the original input. Otherwise, the features of the original image will not be retained in the output.

(3) The role of a decoder is to reconstruct low-level features from feature vectors, which can be accomplished by using a deconvolution layer. Finally, the low-level functions are converted to images in the target domain.

2.3.2. Discriminator Network

The discriminator takes an image as input and predict whether the image is the original image or the output of the generator. The discriminator is composed of multiple convolutional layers. After extracting features from the image, the discriminator can determine whether these features belong to a specific category. The last layer of the discriminator network is the convolutional layer used to generate one-dimensional output.

3. EXPERIMENT AND RESULTS

We compared the BMT-GAN framework with several state-of-the-art brain imaging modality translation approaches. In this section, we will describe the dataset, model performance evaluation, evaluation metrics, qualitative, and quantitative results to demonstrate the effectiveness of the brain imaging modality migration approach.

3.1. Datasets

In our work, we applied the two-dimensional sensor cap image of EEG and the two-dimensional axial slice of the high-resolution T1-weighted image of fMRI. The utilized datasets are illustrated in **Figure 3**.

The EEG and fMRI imaging data from 17 adult volunteers. They were acquired in a Siemens Avanto 1.5T clinical scanner using a self-shielded gradient set with maximum gradient amplitude of $40m^T m^{-1}$, derived from Deligianni et al. (2014).

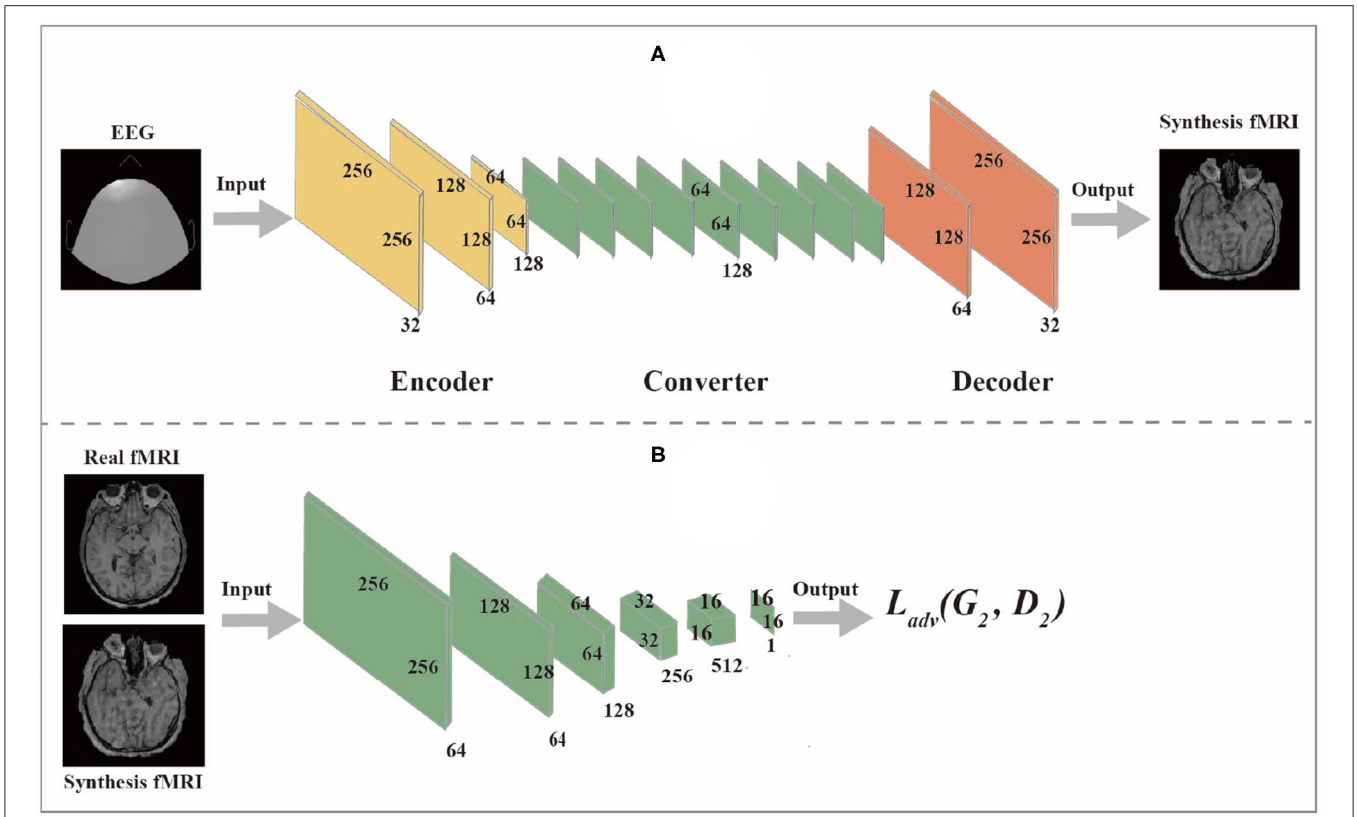


FIGURE 2 | (A) is Generator network, **(B)** is Discriminator network.

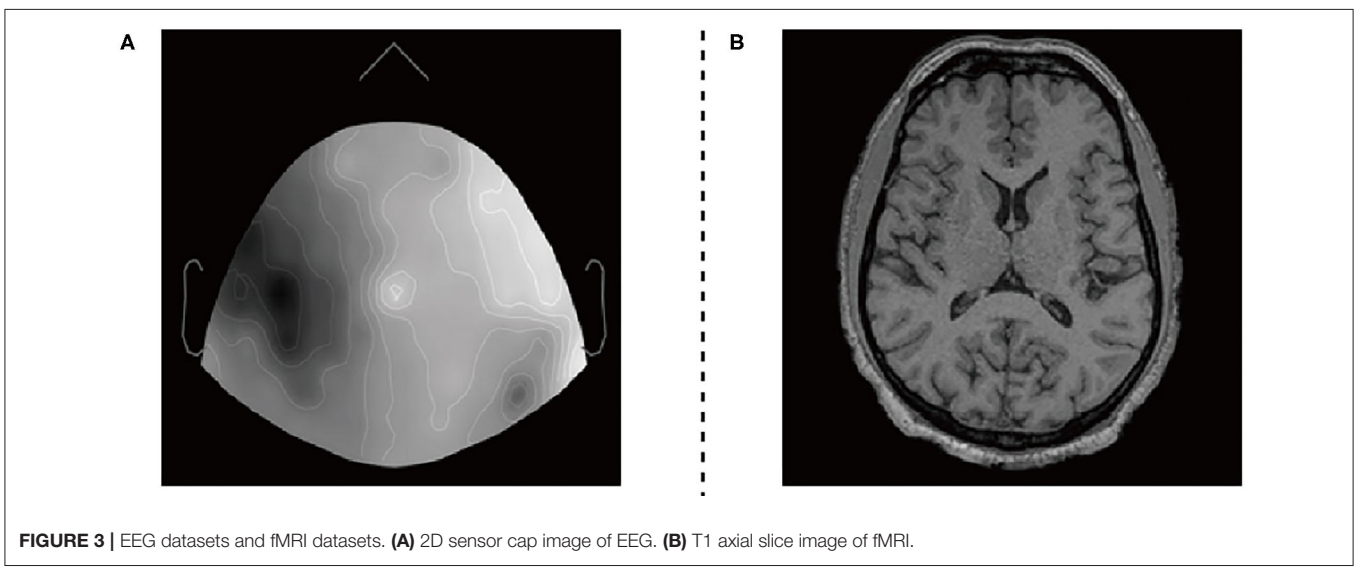


FIGURE 3 | EEG datasets and fMRI datasets. **(A)** 2D sensor cap image of EEG. **(B)** T1 axial slice image of fMRI.

The fMRI imaging data acquisition was based on a $T2*$ -weighted gradient-echo EPI sequence with 300 volumes, and the effective voxel size is $3.3 \times 3.3 \times 4.0mm^3$. Scalp EEG was recorded during the MRI scan using a 64-channel MR-compatible electrode cap (BrainCap MR, Gilching, Germany) at a native frequency of 1,000 Hz.

The two types of imaging data were visualized by using Brainstorm visualization software. Each volunteer involved 35 or more EEG and fMRI two-dimensional axial slices. They have pixels of the same size as $256 \times 256 \times 1$. We divide the datasets into two training sets and two test sets. The training set includes EEG images and fMRI images of 12 subjects, and the test set

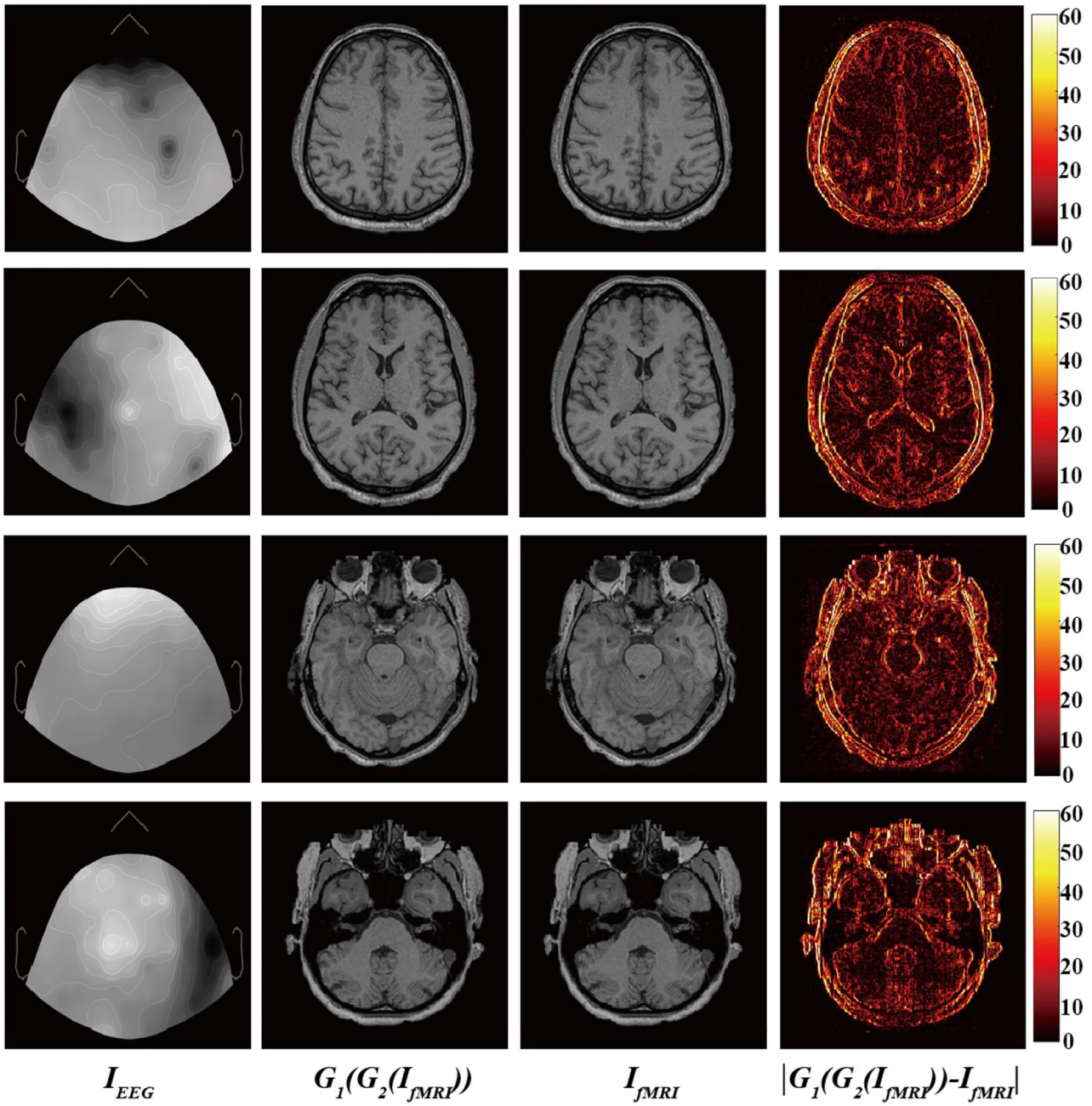


FIGURE 4 | From **(left)** to **(right)**: Input EEG images, output transfer fMRI images, reference fMRI images, and absolute error map between the output translated fMRI and the reference fMRI images.

includes EEG images and fMRI images of the remaining subjects. Secondly, we used the Intel(R) Core(TM) i7-9700 CPU processor for $2e^5$ iterations. The entire training process takes about 40 h.

Finally, **Figure 4** shows the input EEG image, the output fMRI image after modality transfer, the target fMRI image, and the absolute difference map between the output fMRI and the reference fMRI image.

It is worth mentioning that we narrowed the original difference range to [0,60] to make the absolute difference clearer and easier to analyze. Observed from local areas, there are still differences in some local areas. The biggest difference appears in the area of bone structure, while the smallest difference appears in the soft tissue area of brain tissue. This part may be caused by the misregistration between EEG and fMRI images and the

large difference in image structure. From a holistic view, BMT-GAN has successfully learned different anatomical structures in fMRI images (such as bone, gyrus, and soft tissue of brain tissue) in EEG images, and learned to distinguish different anatomical structures with similar pixel intensity. In addition, BMT-GAN also implements the translation of EEG images to fMRI images and ensures the global consistency of the output fMRI images to the reference fMRI images.

3.2. State-of-the-Art Works

We compared BMT-GAN with the following methods:

(1) UNIT (Wang T.-C. et al., 2018): A method of using latent shared space for image transfer.

(2) CGAN (Mirza and Osindero, 2014; Bayramoglu et al., 2017): A method to generate the specified image. It adds additional information, such as labels, to the input of the original GAN generator and discriminator.

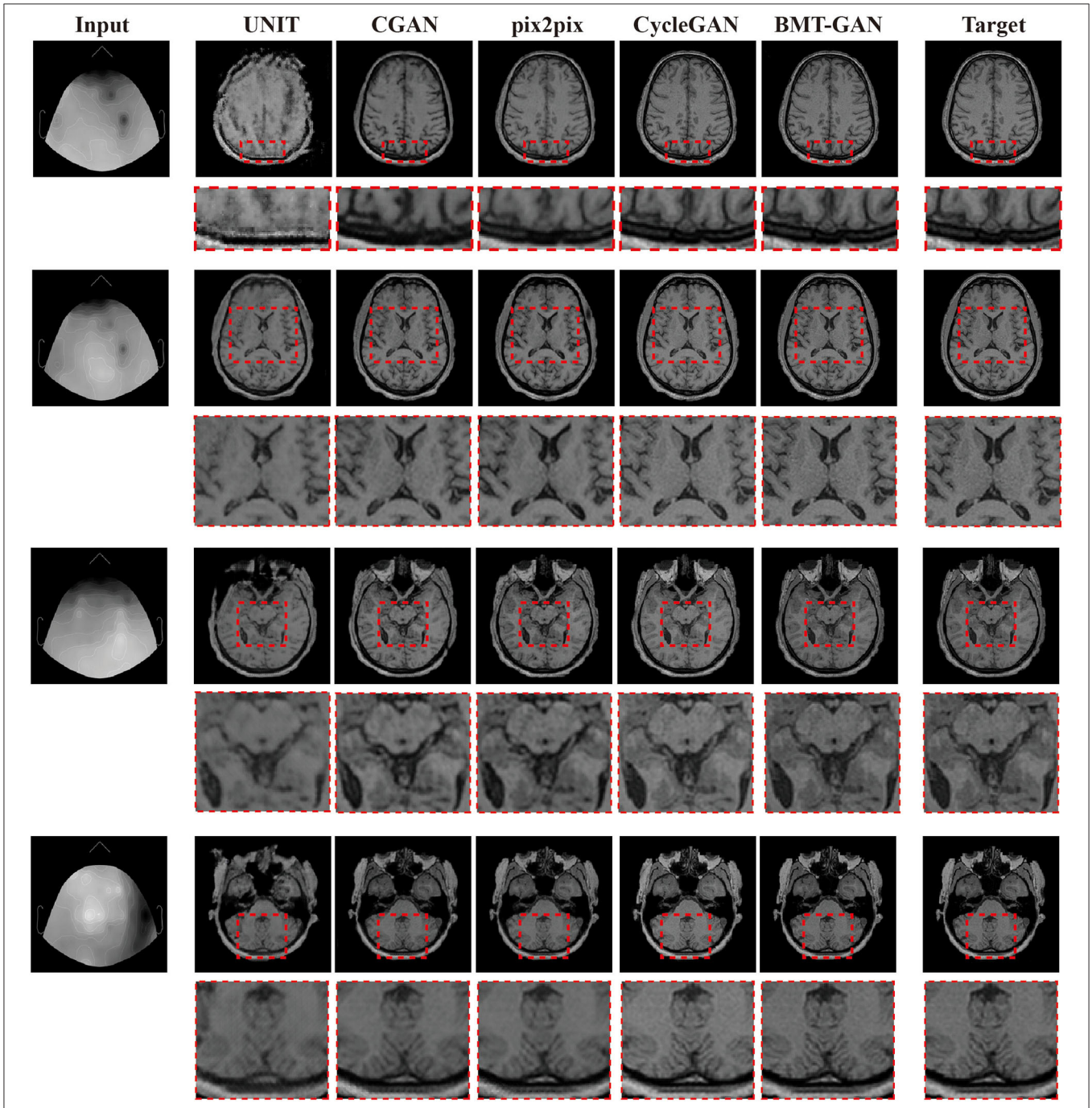


FIGURE 5 | Qualitative comparisons between the BMT-GAN framework and other brain imaging modality transfer technologies. Describes the task of converting EEG to fMRI.

TABLE 1 | Quantitative comparison of brain imaging modality transfer technologies.

Method	MSE	PSNR	SSIM	VIF	IFC
UNIT	1.0230e+03	18.0322	0.5705	0.0465	0.2733
CGAN	340.4902	22.8098	0.6783	0.1614	0.9207
pix2pix	241.0907	24.3090	0.8511	0.3392	2.3423
CycleGAN	143.9875	26.5476	0.8092	0.3323	2.1211
BMT-GAN	128.6233	27.0376	0.8627	0.3575	2.4794

(3) pix2pix (Choi and Lee, 2018; Olut et al., 2018): A method uses a combination of L1 distance and antagonistic loss to train paired data.

(4) CycleGAN (Dar et al., 2019): This method learns two mappings, and introduces two Cycle-consistent loss and adversarial loss to regularize the mapping.

3.3. Evaluation Metrics

We use three commonly used image quality evaluation metrics: Mean-square error (MSE) (Chai and Draxler, 2014), Peak signal-to-noise ratio (PSNR) (De Boer et al., 2003), and Structural similarity (SSIM) (Wang et al., 2004). Their definitions are as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |(I_{fMRI}(i) - G_1(I_{EEG}(i)))| \tag{7}$$

Where it is assumed that the number of 2D slices referred to in the fMRI image is N, i is the index of the image slices, and MAE is the measurement of the average distance between each pixel of the composite image and the real image. The PSNR formula is as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (I_{fMRI}(i) - G_1(I_{EEG}(i)))^2 \tag{8}$$

$$PSNR = 10 * log_{10}\left(\frac{MAX^2_{I_{fMRI}}}{MSE}\right) \tag{9}$$

Where MAX is the maximum intensity value of the original image and the composite image. Since each pixel of the resulting composite image and the original image is represented by an 8-bit binary, $MAX = 255$. Both PSNR and MAE can be used to compare the performance of different models over a unified dataset, but they are limited only to image alignment. Therefore, we also need the structural similarity (SSIM) index, which can not only measure the pixel brightness difference of the image but also evaluate the contrast and structure of the image. Its formula is defined as:

$$\begin{cases} L(x,y)_i &= \frac{2\mu_{i,x}\mu_{i,y}+C_1}{\mu_{i,x}^2+\mu_{i,y}^2+C_1} \\ C(x,y)_i &= \frac{2\sigma_{i,x}\sigma_{i,y}+C_2}{\sigma_{i,x}^2+\sigma_{i,y}^2+C_2} \\ S(x,y)_i &= \frac{\sigma_{i,xy}+C_3}{\sigma_{i,x}\sigma_{i,y}+C_3} \end{cases} \tag{10}$$

$$SSIM(x,y)_i = [L(x,y)_i^{\alpha} * C(x,y)_i^{\beta} * S(x,y)_i^{\gamma}] \tag{11}$$

Set α, β, γ to 1, you can get:

$$SSIM(x,y)_i = \frac{(2\mu_{i,x}\mu_{i,y} + C_1)(2\sigma_{i,x}\sigma_{i,y} + C_2)}{(\mu_{i,x}^2 + \mu_{i,y}^2 + C_1)(\sigma_{i,x}^2 + \sigma_{i,y}^2 + C_2)} \tag{12}$$

Where, $\mu_{i,x}\mu_{i,y}$ are the mean value of the synthetic image and the real image, respectively, and $\sigma_{i,x}\sigma_{i,y}$ are the synthetic image and the real image respectively The variance of the image, $\sigma_{i,xy}$ is the covariance between the synthetic image and the real image, C_1, C_2, C_3 are constants to avoid system errors caused by a denominator of 0.

The above three evaluation metrics measure the quality of the image from a statistical perspective by calculating the pixel error between the output image and the target image. These three evaluation metrics are relatively simple and easy to implement, but they do not take into account the local visual factors of human eyes, so there is no way to grasp the local quality of the image. Recent research has shown that Visual Information Fidelity (VIF) (Sheikh and Bovik, 2006) and Information Fidelity Criterion (IFC) (Sheikh et al., 2006) are widely used to evaluate image quality. They extend the link between the image and the human eyes in terms of fidelity of information. Therefore, we introduce these two valuation metrics, which make image similarity measurement more effective.

3.4. Result Analysis

The proposed BMT-GAN framework in this paper is experimentally verified. Compared with the traditional pix2pix framework, CGAN framework in the supervised field, and the famous CycleGAN, UNIT frameworks in the unsupervised field to prove the superiority of our proposed framework. Through the final five evaluation metrics and convergence performance to compare the similarity of pixel distribution histogram indicators to prove the feasibility and efficiency of BMT-GAN.

3.4.1. Prediction Performance

The experiment results are obtained by using target fMRI and output fMRI images as main examples. During training, the training data are completely consistent, and the same input is used for verification. The prediction results are shown in **Figure 5**. The UNIT framework completes the image transformation between the two domains, but distortion occurs in some areas of the image resulting in the blurred output of the translated image. Qualitatively, pix2pix and CGAN frameworks have also successfully achieved image conversion between two domains. It shows better performance than UNIT but lacks

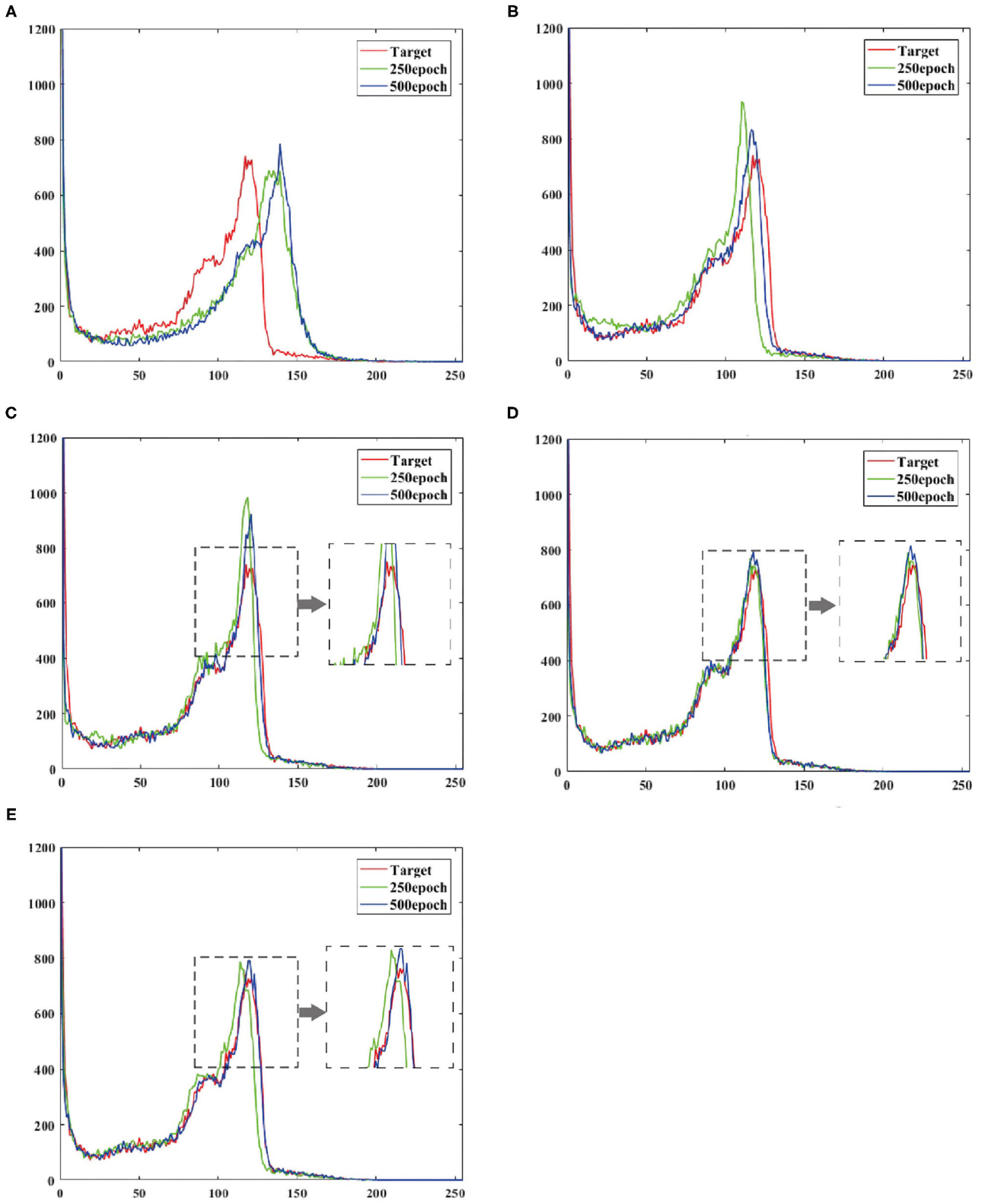


FIGURE 6 | Shows the output reconstruction and the pixel distribution histogram of the target image at the 250th epoch, and the output reconstruction and the pixel distribution histogram of the target image at the 500th epoch. **(A)** UNIT, **(B)** CGAN, **(C)** pix2pix, **(D)** CycleGAN, **(E)** BMT-GAN.

high resolution. The BMT-GAN framework and CycleGAN framework in this paper both have good results. However, the proposed BMT-GAN framework yields results that are closer to the ground truth, more closely connected between output, and input, twisted less and results not only on the style transformation is more apparent. The results show that the framework proposed in this paper has a good application value in the field of image translation.

In order to more accurately verify the performance of the proposed model. We introduce MSE, PSNR, SSIM, and other property price indicators to measure the output results through these numerical scoring standards. The specific results are shown in **Table 1**.

Table 1 shows the MSE, PSNR, and other indicators values computed based on the predicted images with respect to the ground truth. Specifically, BMT-GAN again demonstrates the best performance among all the compared methods. BMT-GAN improves the state-of-the-art PSNR/SSIM performance from 26.5476/0.8092 given by CycleGAN to 27.0376/0.8627. However, as analyzed before, UNIT combined with traditional GAN networks and variational autoencoders to improve the output results, although very good results can be achieved, because two sets of GAN networks are trained at the same time, each step of training requires parameter updates for the two sets of GAN networks, resulting in higher training costs Large, the convergence rate is slower.

TABLE 2 | Quantitative comparison of brain imaging modality transfer technologies.

Method	UNIT	CGAN	pix2pix	CycleGAN	BMT-GAN
EM	0.0438	0.0281	0.0255	0.0214	0.0138
CS	9.7175	2.7339	1.5096	1.3028	0.5531

3.4.2. Convergence Performance

To verify the effect of convergence speed on the model, we divided the training period of the model into 250 epochs and 500 epochs, and obtained the pixel distribution histogram of the reconstructed image and the target image at different epochs. Observe the distance between the pixel distribution histogram of different epochs and the target pixel distribution histogram. The size of the distance indicates how fast the model converges, as shown in **Figure 6**.

At the 250th epoch, we can see that the UNIT model has the slowest convergence speed, and the CycleGAN model has the fastest convergence speed. But because the histogram of the 500 epoch does not give clear results. Therefore, we use two commonly used evaluation indicators: Euclidean metric (EM) and cosine similarity (CS) to compare the similarity between the output reconstructed image of the five models and the pixel distribution histogram of the target image.

Table 2 and **Figure 7** show that the model proposed in this paper is closest to the target value. Compared with other models, the BMT-GAN framework proposed in this paper can more effectively improve the performance of image translation tasks and enhance the relationship between output and input.

4. CONCLUSION

This paper proposes a novel brain imaging modality transfer framework, namely BMT-GAN. The proposed BMT-GAN framework introduces the adversarial loss of discriminator network, dual-period consistent loss of Unpaired training data

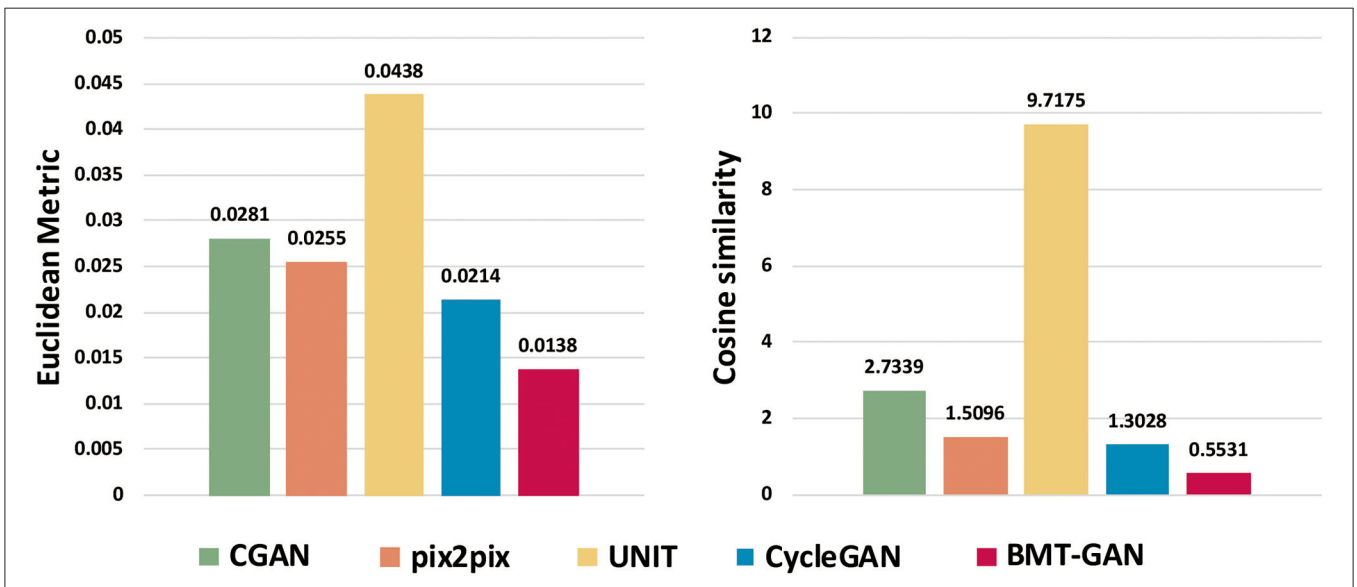


FIGURE 7 | The output reconstruction of five different brain imaging modal transfer methods and the histogram of the Euclidean distance value and cosine similarity value of the pixel distribution histogram of the target image.

and the feature-based non-adversarial loss combination to realize the modality transformation from EEG image to fMRI image. Specifically, the traditional modality transfer methods of brain imaging are based on pixel loss, which usually leads to blurring results. In the field of medical images, small structures can significantly change the diagnostic information of images. Therefore, to capture the difference between the high-frequency components in the image, we introduce non-adversarial perception and style loss based on CycleGAN. This application can be said to be novel. It makes full use of the complementary information of different modality images and can synthesize fMRI images with better tissue contrast and anatomical details, which improves the diagnostic value of brain imaging technology in clinical medicine.

The qualitative and quantitative results show that the performance of BMT-GAN is superior to the existing methods of brain imaging modality transfer. Specifically, the output fMRI images were closer to the target, and the lowest MSE (128.6233), the highest PSNR (27.0367), and the highest SSIM (0.8627) were obtained. However, the deviation between EEG images and reference fMRI images may have a significant impact on quantitative evaluation. Although quantitative measurement is a standard to evaluate the performance of a method, in this case, the numerical difference in quantitative evaluation cannot correctly represent the quality difference. In future work, the accuracy of synthetic fMRI images based on medical expert perception studies will be evaluated.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

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

FUNDING

This research was supported by the National Natural Science Foundation of China (61773244), the General Project of Shandong Provincial Natural Science Foundation (ZR2017MF049), and the Research on University Smart Energy Management System Based on 6LoWPAN (NGII20180609).

REFERENCES

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., et al. (2020). Medgan: Medical image translation using gans. *Comput. Med. Imaging Graph.* 79:101684. doi: 10.1016/j.compmedimag.2019.101684

Bayramoglu, N., Kaakinen, M., Eklund, L., and Heikkila, J. (2017). "Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Venice), 64–71. doi: 10.1109/ICCVW.2017.15

Chai, T., and Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?-arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 7, 1247–1250. doi: 10.5194/gmd-7-1247-2014

Choi, H., and Lee, D. S. (2018). Generation of structural MR images from amyloid pet: application to mr-less quantification. *J. Nucl. Med.* 59, 1111–1117. doi: 10.2967/jnumed.117.199414

Cordier, N., Delingette, H., Lê, M., and Ayache, N. (2016). Extended modality propagation: Image synthesis of pathological cases. *IEEE Trans. Med. Imag.* 35, 2598–2608. doi: 10.1109/TMI.2016.2589760

Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., and Çukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans. Med. Imag.* 38, 2375–2388. doi: 10.1109/TMI.2019.2901750

De Boer, J. F., Cense, B., Park, B. H., Pierce, M. C., Tearney, G. J., and Bouma, B. E. (2003). Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Opt. Lett.* 28, 2067–2069. doi: 10.1364/OL.28.002067

Deligianni, F., Centeno, M., Carmichael, D. W., and Clayden, J. D. (2014). Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. *Front. Neurosci.* 8:258. doi: 10.3389/fnins.2014.00258

Denton, E. L., Chintala, S., szlam, a., and Fergus, R. (2015). "Deep generative image models using a laplacian pyramid of adversarial networks," in *Advances in Neural Information Processing Systems* (Montreal, QC), 1486–1494.

Farnia, F., and Tse, D. (2018). "A convex duality framework for gans," in *Advances in Neural Information Processing Systems, Vol. 31* (Montreal, QC: Curran Associates, Inc.), 5248–5258.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems, Vol. 27* (Cambridge), 2672–2680.

Han, X. (2017). MR-based synthetic ct generation using a deep convolutional neural network method. *Med. Phys.* 44, 1408–1419. doi: 10.1002/mp.12155

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI). doi: 10.1109/CVPR.2017.632

Jin, C.-B., Kim, H., Liu, M., Han, I. H., Lee, J. I., Lee, J. H., et al. (2019). Dc2anet: Generating lumbar spine MR images from CT scan data based on semi-supervised learning. *Appl. Sci.* 9:2521. doi: 10.3390/app9122521

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision-ECCV 2016* (Amsterdam), 694–711. doi: 10.1007/978-3-319-46475-6_43

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., et al. (2014). "Deep learning based imaging data completion for improved brain disease diagnosis," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2014* (Boston, MA), 305–312. doi: 10.1007/978-3-319-10443-0_39

Menon, R. S., Gati, J. S., Goodyear, B. G., Luknowsky, D. C., and Thomas, C. G. (1998). Spatial and temporal resolution of functional magnetic resonance imaging. *Biochem. Cell Biol.* 76, 560–571. doi: 10.1139/o98-051

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv [Preprint]. arXiv:1411.1784.*

Olut, S., Sahin, Y. H., Demir, U., and Unal, G. (2018). "Generative adversarial training for MRA image synthesis using multi-contrast MRI," in *International Workshop on Predictive Intelligence in Medicine* (Granada), 147–154. doi: 10.1007/978-3-030-00320-3_18

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., et al. (2016). "Improved techniques for training gans," in *Advances in Neural Information Processing Systems, Vol. 29* (Barcelona), 2234–2242.

Sheikh, H. R., and Bovik, A. C. (2006). Image information and visual quality. *IEEE Trans. Image Process. A Publ. IEEE Signal Process. Soc.* 15, 430–444. doi: 10.1109/TIP.2005.859378

Sheikh, H. R., Member, IEEE, Bovik, A. C., and Fellow (2006). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process*. 14, 2117–2128. doi: 10.1109/TIP.2005.859389

Tang, H., Xu, D., Sebe, N., and Yan, Y. (2019). "Attention-guided generative adversarial networks for unsupervised image-to-image translation," in *2019 International Joint Conference on Neural Networks (IJCNN)* (Budapest), 1–8. doi: 10.1109/IJCNN.2019.8851881

Wang, C., Xu, C., Wang, C., and Tao, D. (2018). Perceptual adversarial networks for image-to-image transformation. *IEEE Trans. Image Process*. 27, 4066–4079. doi: 10.1109/TIP.2018.2836316

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018). "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/CVPR.2018.00917

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process*. 13, 600–612. doi: 10.1109/TIP.2003.819861

Welander, P., Karlsson, S., and Eklund, A. (2018). Generative adversarial networks for image-to-image translation on multi-contrast MR images - A comparison of cyclegan and UNIT. *arXiv [Preprint]. arXiv:1806.07777*.

Wolterink, J. M., Dinkla, A. M., Savenije, M. H. F., Seevinck, P. R., van den Berg, C. A. T., and Isgum, I. (2017). "Deep MR to CT synthesis using unpaired data," in *Simulation and Synthesis in Medical Imaging*, eds S. A. Tsaftaris, A. Gooya, A. F. Frangi, J. L. Prince (Cham: Springer International Publishing), 14–23. doi: 10.1007/978-3-319-68127-6_2

Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E.-C., Xu, Y., et al. (2018). MRI image-to-image translation for cross-modality image registration and segmentation. *arXiv [Preprint]. arXiv:1801.06940*.

Ye, D. H., Zikic, D., Glocker, B., Criminisi, A., and Konukoglu, E. (2013). "Modality propagation: Coherent synthesis of subject-specific scans with data-driven regularization," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013*, eds. K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab (Berlin: Heidelberg), 606–613. doi: 10.1007/978-3-642-40811-3_76

Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: a review. *Med. Image Anal.* 58:101552. doi: 10.1016/j.media.2019.101552

Zhao, J. J., Mathieu, M., and LeCun, Y. (2016). Energy-based generative adversarial network. *arXiv [Preprint]. arXiv:1609.03126*.

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 17 March 2021
doi: 10.3389/fnins.2021.651920

Quantification of Cognitive Function in Alzheimer's Disease Based on Deep Learning

Yanxian He[1,2†], Jun Wu[2,3†], Li Zhou[2,4], Yi Chen[1,2], Fang Li[2,5] and Hongjin Qian[1,2]*

[1] One Departments of Cadre Ward, General Hospital of Southern Theater Command, PLA, Guangzhou, China, [2] Branch of National Clinical Research Center for Geriatric Diseases, Chinese PLA General Hospital, Guangzhou, China, [3] Three Departments of Cadre Ward, General Hospital of Southern Theater Command, PLA, Guangzhou, China, [4] Two Departments of Cadre Ward, General Hospital of Southern Theater Command, PLA, Guangzhou, China, [5] Four Departments of Cadre Ward, General Hospital of Southern Theater Command, PLA, Guangzhou, China

Alzheimer disease (AD) is mainly manifested as insidious onset, chronic progressive cognitive decline and non-cognitive neuropsychiatric symptoms, which seriously affects the quality of life of the elderly and causes a very large burden on society and families. This paper uses graph theory to analyze the constructed brain network, and extracts the node degree, node efficiency, and node betweenness centrality parameters of the two modal brain networks. The T test method is used to analyze the difference of graph theory parameters between normal people and AD patients, and brain regions with significant differences in graph theory parameters are selected as brain network features. By analyzing the calculation principles of the conventional convolutional layer and the depth separable convolution unit, the computational complexity of them is compared. The depth separable convolution unit decomposes the traditional convolution process into spatial convolution for feature extraction and point convolution for feature combination, which greatly reduces the number of multiplication and addition operations in the convolution process, while still being able to obtain comparisons. Aiming at the special convolution structure of the depth separable convolution unit, this paper proposes a channel pruning method based on the convolution structure and explains its pruning process. Multimodal neuroimaging can provide complete information for the quantification of Alzheimer's disease. This paper proposes a cascaded three-dimensional neural network framework based on single-modal and multi-modal images, using MRI and PET images to distinguish AD and MCI from normal samples. Multiple three-dimensional CNN networks are used to extract recognizable information in local image blocks. The high-level two-dimensional CNN network fuses multi-modal features and selects the features of discriminative regions to perform quantitative predictions on samples. The algorithm proposed in this paper can automatically extract and fuse the features of multi-modality and multi-regions layer by layer, and the visual analysis results show that the abnormally changed regions affected by Alzheimer's disease provide important information for clinical quantification.

Keywords: Alzheimer's disease, quantification of cognitive function, deep separable convolution, channel pruning, convolutional neural network

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Xiaoqing Gu,
Changzhou University, China
Shan Zhong,
Changshu Institute of Technology, China

****Correspondence:***
Hongjin Qian
drqianhj@126.com;
ld88654191@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *11 January 2021*
Accepted: *08 February 2021*
Published: *17 March 2021*

Citation:
He Y, Wu J, Zhou L, Chen Y, Li F and Qian H (2021) Quantification of Cognitive Function in Alzheimer's Disease Based on Deep Learning. Front. Neurosci. 15:651920. doi: 10.3389/fnins.2021.651920

INTRODUCTION

Alzheimer disease (AD) is a neurodegenerative disease in the brain. It is one of the most common types of dementia, accounting for about 60–80% of the total number of dementia patients (Lee et al., 2019; Spasov et al., 2019). AD usually has a chronic or progressive nature, and a variety of high-order cortical dysfunctions appear, manifested by symptoms such as cognitive decline, reduced judgment and memory loss, and ultimately lead to the loss of independent living ability. Many AD patients need to rely on the help of others to maintain a normal life, which brings a huge burden to caregivers (Martinez-Murcia et al., 2019; Noor et al., 2020). These burdens include various pressures on social life, psychological aspects, physical activities, and economic levels. AD has become a common problem faced by the whole world. To overcome the pathogenesis of AD, it is urgent to carry out early quantification and treatment. Mild Cognitive Impairment (MCI) is the early stage of AD. The annual conversion rate of MCI to AD is as high as 10–15%, while the annual conversion rate of healthy people to AD is only 1–2% (Ju et al., 2017). Certain cognitive training and rehabilitation can delay the development of AD, and some can return to normal. However, once it enters the AD stage, no effective therapeutic drugs have been developed clinically, and this process is irreversible (Bi et al., 2020a; El-Sappagh et al., 2020).

In recent years, the research on the combination of artificial intelligence technology and medical big data has achieved a lot of research success in the medical field (Li et al., 2019). Artificial intelligence technology can help doctors and patients predict more potential diseases. At the same time, medical big data also provides a good foundation and platform for the application of artificial intelligence technology. Among them, deep learning is a member of the artificial intelligence army. The core concept of deep learning is to simulate the multi-level information processing method of the brain, and interpret the input information in layers to obtain a series of characteristics of the input information (Luo et al., 2017; Shi et al., 2017). In the medical field, deep learning can not only perform image analysis and intelligent quantification of diseases, but also improve the efficiency of medical data collection and processing, thereby improving the accuracy of doctors in diagnosis and treatment of diseases, so that patients can get more timely, more complete, and more accurate treatment (Chen et al., 2019).

In actual clinical quantification, doctors often need to analyze multiple modalities of image data, and integrate multiple quantitative information, combined with experience knowledge, in order to make an objective judgment on the patient's condition. This paper uses graph theory to extract brain network features and verify the effectiveness of the features. In this paper, two kinds of brain networks are established, and three graph theory parameters of node efficiency, node degree, and node betweenness centrality are extracted respectively. Through significant difference analysis, graph theory parameters with obvious differences between normal people and AD patients are taken as brain network characteristics. Specifically, the technical contributions of this article can be summarized as follows:

First: This article analyzes the computational complexity in the convolutional layer, introduces several special convolution structures, and focuses on the depth separable convolution structure, and based on the convolution unit, a new channel pruning is developed. We elaborated the channel pruning process for a single convolution unit, analyzed the compression effect of this method on the convolution unit, and then introduced the pruning process for the entire network. The key issues of channel selection, pruning ratio selection and model performance recovery in the overall pruning process are discussed. The channel selection was carried out according to the APo Z channel importance evaluation criteria, and the channel pruning was carried out on Mobile Net.

Second: We try to use multi-modal brain imaging data, such as MRI images. Through non-invasive imaging technology, clear tissue structure of the patient's brain can be obtained, PET images and changes in the function of various brain tissues can be obtained through changes in glucose metabolism. We combine the two modal data for comprehensive analysis, which can improve the specificity and sensitivity of AD and MCI quantification, effectively prevent misdiagnosis and missed diagnosis, and improve the credibility of computer-aided quantification.

Third: In this paper, 3D-CNNs are used to extract the features of local 3D image blocks, and the feature output of the intermediate convolutional layer corresponding to the MRI and PET images at the same location is taken as the input of the feature fusion network, and the local features at different locations are stitched together, and then we use the trained network to make the final quantification of the sample. In the experiment based on multi-modal data, the quantization accuracy of AD and NC based on cascaded 3D-CNNs reached the ideal level.

The rest of this article is organized as follows. Section 2 discusses related work. Section 3 carried out the extraction and analysis of the parameter features of Alzheimer's disease brain network graph theory. Section 4 designs a channel pruning algorithm based on efficient convolution unit. Section 5 presents the results and analysis of cognitive function quantitative experiments. Section 6 summarizes the full text.

RELATED WORK

AD is essentially a disease that continues to deteriorate and is incurable, and the direct cause is still unknown. The main factors that cause Alzheimer's are genetic, neurotransmitter, immune and environmental factors (Lu et al., 2018a; Jain et al., 2019; Raza et al., 2019). Although the pathological changes of AD begin to appear very early, its typical clinical symptoms do not show up until later. According to estimates by the Alzheimer's Association, if there is no major breakthrough in AD prevention, by 2050, the number of AD patients worldwide is expected to exceed 155 million (Zeng et al., 2018). Studies have found that about 12% of MCI patients convert to AD every year, while the annual probability of converting to AD for elderly people with normal cognitive function is only 1–2% (Oh et al., 2019).

The high conversion rate of MCI to AD on the one hand shows that MCI is a high-risk group of AD, on the other hand, it also reflects the importance of early quantification of AD (Bi et al., 2020b). Therefore, if the patients with MCI can be screened out based on the early quantitative examination reports of the patients and their disease development can be predicted, preventive intervention can be carried out on the patients and standardized drug treatments can be given to the patients, thereby preventing or delaying the occurrence of AD (Choi et al., 2018; Lu et al., 2018b).

When doctors are faced with huge medical pathology data, the workload of quantifying AD is too complicated and there are some subjective predictions. For example, MRI images require doctors to go through the naked eye, linear measurement, area measurement, volume measurement, and MRI value measurement. Therefore, the accuracy and efficiency of AD quantification results may be further improved. At present, most researches use simple data analysis techniques and machine learning techniques to make quantitative predictions on MRI images, and on the other hand, researches are based on a single AD neuroimaging data, molecular biology data or genetic examination data. The use of deep learning technology to analyze MRI image performance text reports and multiple AD clinical examination data has not been studied in depth. In response to the above problems, related scholars proposed a new quantitative model and predictive model based on deep learning, designed and implemented an Alzheimer's assisted quantitative medical system, which separately reported on MRI images. And multiple AD clinical examination data quantification can assist doctors in quantifying and predicting the development of the patient's condition, providing more treatment time for preventing or delaying AD.

The quantification of Alzheimer's disease needs to be combined with the patient's medical history, family history, neuropsychological evaluation and other examinations, and the cause of the quantification needs to be based on clinical manifestations, biomarkers and structural images (Tang et al., 2019). Nowadays, domestic and foreign researches on quantitative prediction methods for Alzheimer's disease are mostly based on MRI images in structured images, and then use different research methods for research, as well as using biomarkers and neuropsychological examinations (Wang et al., 2019).

Structured images include Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). The early lesions of AD mainly involve the hippocampus, and CT is difficult to accurately display the structure of the hippocampus. Therefore, the role of CT in distinguishing and quantifying AD is limited. MRI shows that the abnormal changes of dementia are more sensitive than CT, and is recognized as the best imaging method to quantify and display the morphological abnormalities of dementia (Li et al., 2018). Therefore, MRI is more popular in quantitative research on the development of AD. Relevant scholars have investigated the gray matter and differences in MRI images, using posture morphology and statistical analysis methods to achieve the quantification of normal, mild cognitive impairment, and Alzheimer's disease (Zeng et al., 2019). Based on Adaboost, the researchers analyzed the hippocampus volume in MRI images and studied the significant differences in the low-frequency amplitude of the three brain regions of normal, mild cognitive impairment, and Alzheimer's, so as to quantify the condition of Alzheimer's disease (Choi et al., 2019). Relevant scholars use principal component analysis to reduce the dimensionality of MRI images, and then use linear regression models to predict the development trend of the MMSE scale for patients with Alzheimer's disease in one year (Gulhare et al., 2017; Vu et al., 2018). However, principal component analysis is used for dimensional reduction techniques. The disadvantage is that the interpretability of the established model is poor.

The method based on the prior knowledge area is based on the prior knowledge obtained by researching AD histological or imaging data. Generally, the features of some important regions can express information with rich discriminative power for AD, and these features can be extracted for quantification. The hippocampus is located in the medial temporal lobe and is one of the few areas where severe structural changes occur in AD (Mazurowski et al., 2019). Therefore, the geometrical characteristics of the hippocampus are often used as effective biomarkers. Research by related scholars has shown that hippocampal texture is better than volume shrinkage measurement in terms of predicting the conversion from MCI to AD based on SVM (Kumar et al., 2017). In fact, the effects of AD also manifest in other brain regions. Related scholars use the shape differential morphology measurements of the left and right amygdala, hippocampus, thalamus, caudate nucleus, putamen, globus pallidus, and lateral ventricle as features, and use LDA to predict AD transformation (Zhang et al., 2019).

The rapid development of deep learning has become a good supplement to traditional machine learning algorithms, and has also provided new means for the quantification and prediction of various neurodegenerative diseases, and has been increasingly used in the field of neuroimaging (Kam et al., 2019). However, deep learning algorithms still have certain limitations in neuroimaging research. The existing quantification methods of AD, MCI, and HC based on deep learning have generally achieved high accuracy in the quantification of AD group vs. HC group, but the accuracy rate is slightly less in AD group vs. MCI group and MCI group vs. HC group. This is due to the large differences in the imaging data of AD and NC patients. In the current quantification of AD, MCI and HC by researchers, the accuracy rate is relatively low. On the one hand, although deep learning models usually do not require pre-defined features and can search for and discover complex structural features that characterize the problem based on data, the feature extraction and analysis technology of neuroimaging is currently relatively mature and will be based on neuroimaging preprocessing (Vieira et al., 2017; Le et al., 2019). The extracted features are used to construct a deep learning model, which is very likely to accelerate the convergence speed of model training, reduce the demand for training data, and improve model performance (Liu et al., 2018). On the other hand, the large data training samples required for deep learning and the large number of parameters required for model adjustment can easily lead to a much higher computational complexity and the amount of data required for training models

than traditional machine learning models (Jha and Kwon, 2017; Huang et al., 2020; Wingate et al., 2020). Therefore, although current neuroimaging databases have been well developed, their scale may limit the performance of deep learning to a certain extent. Therefore, this article uses traditional machine learning methods to analyze and study the quantitative problems of AD, MCI, and HC.

ALZHEIMER'S DISEASE PARAMETER FEATURE EXTRACTION

Multimodal MRI Image Preprocessing

The scanned magnetic resonance images are all images output by the machine. These images have noise and the format is not convenient for computer processing. In order to facilitate the further processing and research of the image, it is necessary to perform the image preprocessing process first. There are many steps in the preprocessing process of the three modalities of magnetic resonance images are similar. We will introduce the preprocessing process of the different modalities one by one. **Figure 1** shows the flow of magnetic resonance image processing for Alzheimer's disease.

First we complete the preprocessing process of T1 weighted image. The first step is format conversion, which converts DICOM files into NIFTI files. The format of the original MRI image is generally DICOM format. DICOM is the unified output format of medical imaging machines. The images scanned in this format are independent images and need to be read with specific software, which is very inconvenient. It is not conducive to continue processing research. NIFTI file is an advanced medical image storage format, which has many advantages, such as convenient calculation and analysis, strong format universality, and centralized collection of corpus. We can continue processing the converted data.

The second step is head movement correction. When a patient is scanning for magnetic resonance imaging, it is inevitable that there will be some slight jitter, which will cause noise and artifacts in the collected images. In order to remove artifacts and noise, it is necessary to perform head movement correction on the scanned T1-weighted image. Spatial standardization is an important step to eliminate individual differences. Due to differences in the patient's head volume, shape, etc., all images are mapped once and mapped to the MNI standard space. This will eliminate the influence of patient individual differences on subsequent experiments and make the results more accurate. The fourth step of smoothing filtering is also mainly to eliminate the noise of the patient during the scanning magnetic resonance.

The last step is the division of brain regions. The brain division of T1-weighted image is not actually a routine preprocessing step. In this study, the T1 weighted image is mainly used as a template for the construction of the brain network using f MRI and DTI data. So here is included in the preprocessing process. This process will use the brain division method to divide the brain into 52 brain regions. The preprocessing process of T1 weighted image is now completed.

The f MRI preprocessing process is implemented on the MATLAB open-source toolkit GRETNA (Graph Theoretical Network Analysis Toolbox). The second preprocessing step of f MRI is to remove the initial data. The initial 5 data of f MRI are unstable because of the unstable blood oxygen signal, so the data obtained by f MRI scan is not reliable and needs to be removed. The third step of preprocessing is time level correction. f MRI scans the odd-numbered slices first and then the even-numbered slices during the scanning process. The scanning time of adjacent slices is very different, and the processing can be continued after time correction.

The process of DTI preprocessing is done on MATLAB's open source toolkit PANDA (Pipeline for Analyzing Brain Diffusion Images). There are only two differences between DTI preprocessing and T1 weighted image preprocessing. The second step of pre-processing is to remove the skull. DTI is a structural image, and the structure of the skull is more obvious. In order to prevent the skull from interfering with the research in the subsequent research process, it was removed in the preprocessing process. The fourth step of pretreatment process is eddy current correction. Due to the frequent switching of the dispersion gradient, the image will produce eddy current distortion. The eddy current correction can remove the eddy current distortion and make the imaging result more accurate. So far, the preprocessing of f MRI, DTI, and T1-weighted images is all over.

Brain Function Connection Network Acquisition

Brain functional connection is obtained from f MRI data, which can reflect the synchronization of functional activities between brain regions by calculating the correlation of time series signals between brain regions. The network formed by the functional connections between all brain regions of the whole brain is called Functional Connectivity Network (FCN).

Through the preprocessing process, the text has obtained the brain template. In this way, the nodes of the brain network can be defined smoothly. In the process of defining the brain function to connect the network edge, the text is defined by the correlation of the mean value of all voxels between the two brain regions. The formula for calculating the time series correlation coefficient between any two brain regions is as follows:

$$c_{a,b}(u) = \frac{Cov_{a,b}(u)}{\sqrt{var(a) \cdot var(b)}} \tag{1}$$

In the above formula, var(a) and var(b) respectively represent the variance of the mean value of all voxel time series in the two brain regions over time, and Cov a b(u) represents the covariance of the time series mean value of the two brain regions. According to the functional magnetic resonance imaging data, the time series of each voxel can be obtained, and the calculation of the time series correlation coefficient between any brain regions can be completed through mathematical calculation.

The obtained correlation coefficient between brain regions is the strength of the functional connection, but only when the correlation coefficient between the two brain regions exceeds



FIGURE 1 | The original image preprocessing process of Alzheimer's disease.

the threshold, can it be determined that there is a functional connection between the two brain regions. The larger the correlation coefficient value, the higher the even strength. The side of the brain function connection network is the above-mentioned functional connection strength. The functional connection strength between any two brain areas of the divided 52 brain areas is calculated, and the acquisition of the brain function connection network is completed. The schematic diagram of the acquisition of the entire brain function connection network is shown in **Figure 2**.

The brain function connection network will be expressed in the form of a mathematical matrix FC. The number of rows and columns of the matrix corresponds to the number of nodes in the brain network. The element values in the matrix represent the magnitude of the functional connection strength between the abscissa brain area and the ordinate brain area. All elements together constitute the weight of the brain function connection network edge.

The white matter structure connection measures the connection of white matter fibers that actually exist between two brain regions. The network formed by the connection of white matter fibers between all brain regions of the whole brain is called DTI Structural Connectivity Network (DTISCN). The white matter structure connection network acquisition is similar to the process of f MRI data acquisition FCN. It also defines the nodes of the network first, and then calculates the edges of the network. The process of defining nodes in the white matter structure connection network is the same as that of FCN. The two imaging methods of the same sample use exactly the same templates, so the white matter structure connection network also has 52 nodes.

The white matter structure connection network will use the anisotropy value FA obtained from DTI data to define the white matter fiber connections between brain regions. The physical meaning of FA value indicates the strength of dispersion, and the number and density of white matter fibers can be reflected by the size of FA value. The method of solving the FA value of each voxel is as follows:

$$FA = \frac{\sqrt{3 \cdot \left[(\eta_1 - \eta'^2) + (\eta_2 - \eta'^2) + (\eta_3 - \eta'^2)\right]}}{\sqrt{2 \cdot \left(\eta_1^2 + \eta_2^2 + \eta_3^2\right)}} \quad (2)$$

Different values of η indicate the degree of dispersion of each voxel in different directions, and η' indicates the average value of the degree of dispersion in different directions. The FA value between any two brain regions can be obtained from the FA value of each voxel of these two brain regions. However, not all the FA values of the brain regions are reasonable and effective, because some FA values cannot describe the white matter fiber connections in the brain regions. Here we will introduce a probabilistic fiber tracking algorithm, which can eliminate invalid FA values and leave valid FA values. The probabilistic neural tracking algorithm calculates the relationship between the gradient information and the anisotropic FA. The threshold of FA is 0.2. If the anisotropy value is lower than 0.2, it is assumed that there are no nerve fibers in the voxel or that nerve fiber disconnection occurs. The threshold of the angle is set

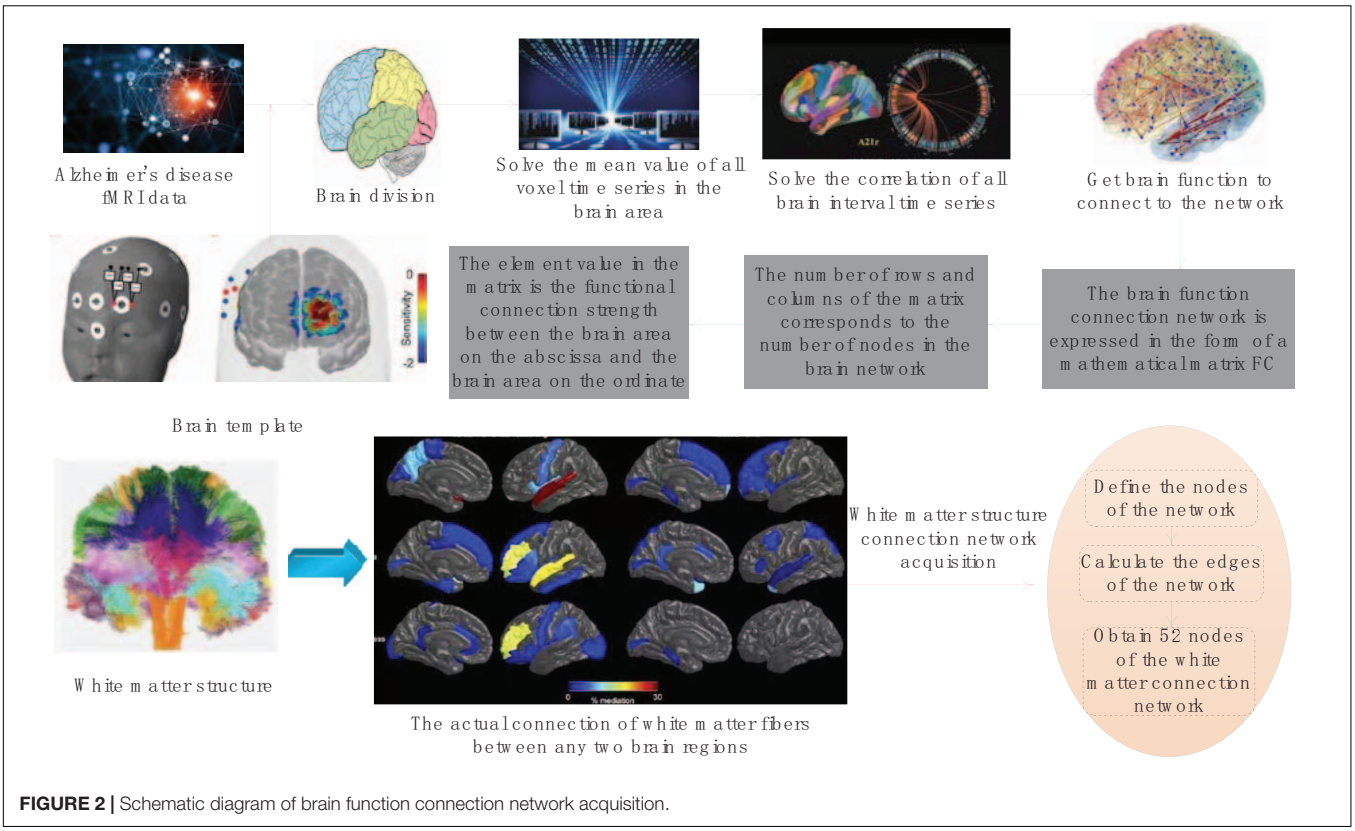


FIGURE 2 | Schematic diagram of brain function connection network acquisition.

to 35°. If the angle exceeds 35°, it is considered the intersection of two nerve fibers instead of one nerve fiber.

In this paper, the white matter structure connection network is obtained with the assistance of MATLAB toolkit PANDA. We can obtain the nodes and edges of the white matter structure connecting network. When the FA value is still confirmed after the fiber tracking between the two brain regions is completed, it is assumed that there is a white matter fiber connection between the two brain regions. The larger the FA value, the stronger the white matter fiber connection strength. Similar to FCN, DTISCN will be obtained by the anisotropic FA matrix between brain regions. The FA matrix can also be visualized for intuitive observation. Similar to the FC matrix, each point of the FA matrix represents the strength of the white matter fiber connection between the abscissa brain area and the ordinate brain area. Because the fiber connection has no directionality, the matrix is a symmetric matrix. Through the visualized graph, it can be seen that the FA value does not exist in many places. One part of the FA value does not exist originally, and the other part is that the FA value is removed after the fiber tracking is completed. The FA matrix can be used to extract the connection network characteristics of the white matter structure.

Extraction of Parameter Features

Graph theory parameters include node degree, node efficiency, and node betweenness centrality. The calculation methods and physical meanings of different parameters are different. The node degree is a description of the importance of a node in the network, specifically expressed as the sum of the weights of all other node edges in the network that have a connection relationship with the node. The larger the node degree, the closer the connection between the node and other nodes, and it also reflects the high importance of the node in the entire network. The calculation formula of node degree of i-node is as follows:

$$D_i = \sum_{j \to N} w_{ij} \tag{3}$$

In the above formula, w_{ij} represents the weight value of the edge between node i and node j, and N represents the node in the network. The value in the connection matrix FC reflects the weight of the brain function connecting the network edge. The anisotropy matrix FA reflects the weight of the white matter structure connected to the network edge. In this study, changes in the degree of nodes can reflect changes in the importance of corresponding brain regions. By comparing normal people with AD patients, changes in the importance of brain regions in AD patients can be found.

Node efficiency, also known as local efficiency, mainly reflects the efficiency of the information transfer process between the node and surrounding nodes. This indicator can not only reflect the efficiency of information flow between the neighbors of the node, but also reflect the degree of optimization of the local network. Therefore, the node efficiency can be expressed as the average of the sum of the reciprocal of the shortest path of each

node in the network composed of the node's neighborhood. The calculation formula for node efficiency of i-node is as follows:

$$E_i = \sum_{j \neq k} \frac{1}{L_{jk}} \frac{1}{N_{G_i}(N_{G_i} - 1)} \tag{4}$$

In the above formula, G_i is a network formed by the neighborhood of node i, and j and k are two nodes in the neighborhood network. L_{jk} is the shortest path between any two nodes in the neighborhood network. According to the FA matrix and the FC matrix, the node efficiency of each node in the two networks of FCN and DTISCN can be calculated respectively. Node efficiency can reflect the efficiency of brain processing information and the ability to resist attacks. The decline in the efficiency of brain nodes will reflect the damage to the brain by diseases.

Node betweenness centrality is also an index used to describe the role and status of a node, which is different from the perspective of node degree description. Betweenness centrality describes the criticality of nodes in the process of network information processing from the perspective of information flow. Node betweenness centrality is defined as the ratio of the number of paths containing the node among all the shortest paths in the entire network to the number of all shortest paths in the entire network. If the betweenness centrality of nodes changes, it means that the shortest path of the whole brain network will change, and the efficiency of the entire network will also change. The formula for calculating betweenness centrality of i-node is as follows:

$$B_i = \sum_{i \neq j \neq k} \frac{\sigma_{jk}(i-1)}{\sigma_{jk}} \tag{5}$$

Node betweenness centrality can also be obtained through FA and FC matrices. Betweenness centrality is a very useful indicator for brain networks, it reflects the size of the brain area in processing information. The change of betweenness centrality in AD patients can reflect the damage and involvement of the brain area.

Graph theory parameters are obtained above, and each graph theory parameter is a characteristic component. But these graph theory parameters can only be used as feature components, and only feature components with significant differences can be called features. Therefore, this article will verify the differences of graph theory parameters, and select the feature components with significant differences as the features of FCN and DTISCN.

In the DTISCN features, the left side of the basal forebrain and the right side of the basal forebrain showed changes in graph theory parameters. It can be inferred that the structure of the basal forebrain has been damaged. Of course, there are many features that are also corroborated with the symptoms of AD patients. I cannot explain them one by one here, but it is certain that there are some abnormalities in the brain function network and white matter structure connection network of AD patients. At the same time, these features are reasonable and effective.

CHANNEL PRUNING ALGORITHM

Complexity of Convolutional Neural Network

In deep neural networks, complexity is divided into computational complexity and space complexity, which have the following effects on the network:

(1) Computational complexity
The training and prediction speed of the model is determined by the computational complexity. The higher the computational complexity, the more time it takes for the forward calculation of the model. The model cannot be tested quickly and cannot be applied in scenarios with high real-time requirements. At the same time, the higher the computational complexity, the more time it takes to train the model. The longer it is, the more it is impossible to verify and improve the model in time.

(2) Space complexity
The number of parameters in the model determines the space complexity of the model. Generally speaking, a larger model has more parameters, and the better the fitting ability of the model. Then the larger the model, the more data is needed to perform the model. The data set in real life is usually not too large, which makes the model more prone to overfitting.

The main operation layer in the convolutional neural network is the convolution layer, and the calculation principle of the convolution layer is as follows:

$$S(i,j) = \sum_m \sum_n K(i-m-1, j-n-1) \cdot I(m,n) \tag{6}$$

Suppose I is the input layer, K is the convolution kernel, the size of I is D1 × D1 × M, D1 is the width and height of the input feature map, M is the dimension of the input feature map, and the convolution kernel size is D2 × D2 × M × N, I get an output feature map with a dimension of D1 × D1 × N through the convolution kernel, so the computational complexity is O(D1 × D1 × D2 × D2 × M × N), and the convolutional layer space complexity is O (D2 × D2 × M × N), the total computational complexity and space complexity of the convolutional neural network is the sum of the computational complexity and space complexity of each convolutional layer.

Some convolution kernels in Alex Net use a size of 5 × 5 or even 7 × 7, which greatly increases the complexity of the network. Starting from the VGG neural network, convolutional neural networks have generally adopted 3 × 3 convolution kernels to obtain features. When cutting the model, since the spatial size of the convolution kernel (3 × 3) is already small, the cutting of the network model usually starts with the number of channels, that is, reducing the value of M or N.

Inception_V1 constructs four parallel convolution/pooling modules of different sizes in an Inception structure, which effectively increases the width of the network, but doing so also causes a surge in the time and space complexity of the network.

The Inception module first uses 1 × 1 convolution to reduce the dimensionality of the input feature map, and performs a weighted feature combination on the features of different channels, and then uses 3 × 3 convolution to simultaneously map the spatial dimension and the channel dimension. It first uses 1 × 1 convolution on the channel correlation to map the input feature map to a space with several dimensions smaller than itself, which is equivalent to multiplying each channel map by a different factor to make a linear combination, using 3 × 3 Convolve these small spaces, and map its space and channel correlation at the same time. The convolutional neural network structure is shown in **Figure 3**.

Using 1 × 1 convolution to reduce dimensionality can reduce the computational complexity by more than 3 times. According to the two-dimensional convolution input and output size relationship, for the same input size, the output of a single 5 × 5 convolution is exactly the same as the output of two 3 × 3 convolution cascades, that is, the receptive field is the same.

Also according to the complexity analysis formula, this replacement can effectively reduce the space and time complexity. Using this convolution structure can use the saved complexity to increase the depth and width of the model, so that the complexity of the model remains unchanged.

Depth Separable Convolution and Pruning

For convolution, the convolution kernel can be regarded as a three-dimensional filter: channel dimension + spatial dimension (corresponding to the width and height of the feature map). The conventional convolution operation is actually to realize the combination of channel correlation and spatial correlation. There is an assumption behind the Inception module: the combination of spatial features of the convolutional layer and the combination of channel features can be performed separately, and better results can be achieved by doing it separately. Depth separable convolution solves traditional convolution into deep convolution and channel convolution. The comparison of ordinary convolution, spatial convolution and channel convolution is shown in **Figure 4**.

The 1 × 1 convolution method of the deep separable convolution structure greatly reduces the amount of calculation in the forward operation process. Not only that, in Mobile Net, about 95% of the multiplication and addition operations come from 1 × 1 convolution (accounting for parameter 95% of the number), the large use of 1 × 1 convolution means that it can be directly implemented using highly optimized matrix multiplication algorithms (such as GEMM), which greatly improves computational efficiency. The method of network pruning has been widely used in convolutional neural network model compression. In early work, network pruning was considered as an effective method to reduce network complexity and reduce overfitting. Pruning the model with the best performance so far reduces the network complexity without loss of accuracy. Generally speaking, the following methods are used for network pruning:

(1) Use ordinary methods to train a complete convolutional neural network. (2) Sort the weight of each layer according to the absolute value of its weight. (3) Select the weight whose absolute value is lower than a certain threshold and remove it. (4) Retrain the network after pruning to achieve the performance before pruning as much as possible.

Usually this method can greatly reduce the number of network parameters in the fully connected layer. The method of network weight pruning can often greatly reduce the amount of network parameters, but weight pruning often has the following disadvantages.

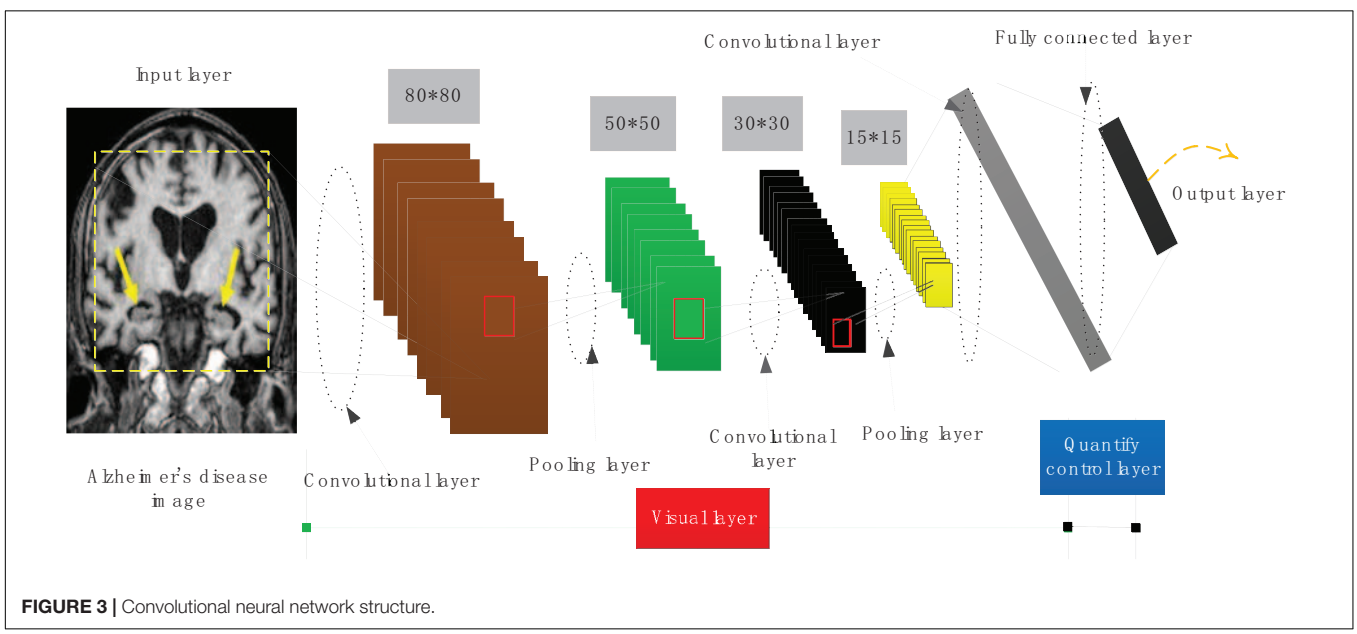


FIGURE 3 | Convolutional neural network structure.

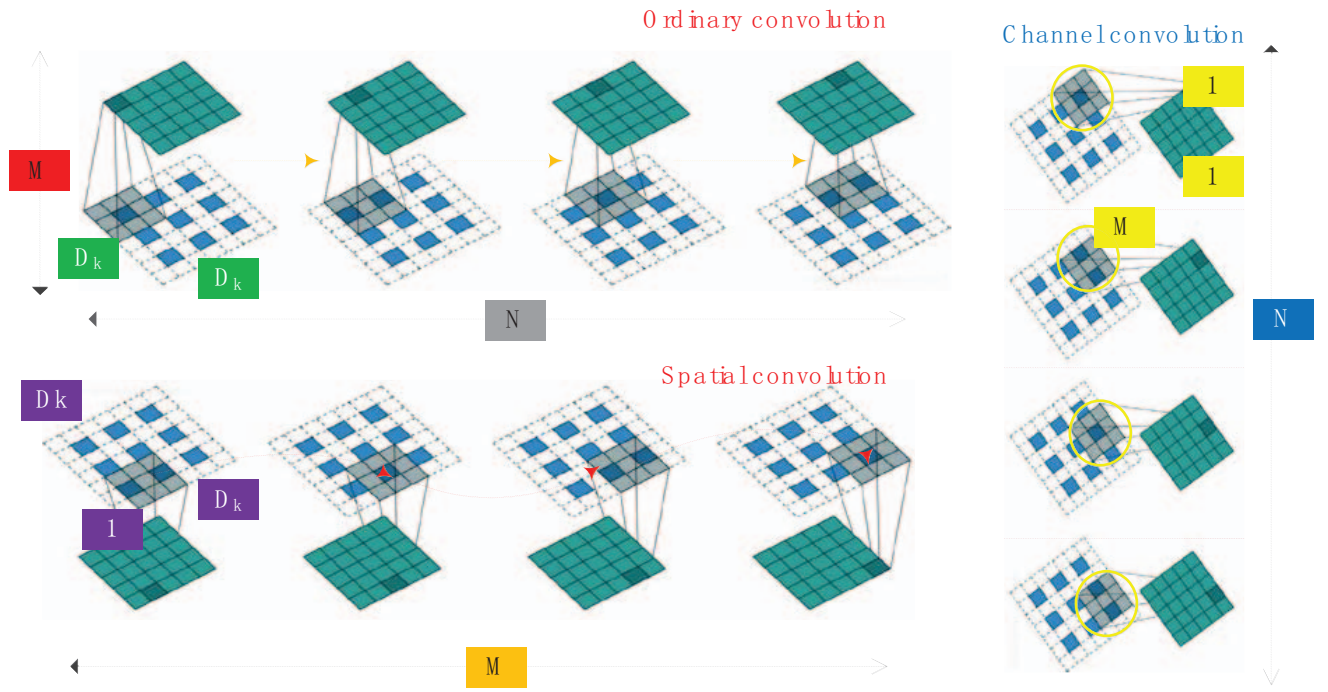


FIGURE 4 | Comparison of ordinary convolution, spatial convolution and channel convolution.

(1) These network pruning methods are only for the fully connected layer. The fully connected layer is often the part with the most redundant parameters. In practical applications, we often abandon the fully connected layer and replace the corresponding part with the average pooling layer. Therefore, in the existing convolutional neural networks, the convolutional layer often accounts for most of the calculation and is the most time-consuming part. In general, the above algorithms can achieve faster speeds or less storage capacity, but they rarely achieve significant acceleration while compressing the entire network.
(2) Weight pruning is easy to produce sparse connections, and the computational efficiency of sparse neural network structure is not as good as that of ordinary tightly connected neural networks.

Channel pruning is another weighted pruning method. Unlike the pruning method that removes a single neuron connection in neuron pruning, channel pruning removes less important channels in the entire convolutional layer. Each filter corresponds to a channel of the activation layer, and the expressive ability of each channel filter is closely related to its corresponding activation layer. A simple strategy to calculate the expression ability of the filter is to calculate the average percentage of zero activation value (Average Percentage of Zeros, APo Z) for each channel of the activation layer. The higher the APo Z, the lower the importance of the filter, and the lower it should be removed after a certain threshold.

Channel Convolution

Currently, for the problem of network pruning, there is no suitable benchmark network architecture as a criterion for judging the performance of pruning. At present, the most commonly used convolutional neural network structures in network pruning, such as Alex Net, Goog Le Net, Res Net, etc., these models are effective in image quantization tasks, but they are in order to achieve the best performance and performance in the Image Net competition. The extreme accuracy rate increases, and the design parameters are seriously excessive, so these convolutional neural networks can easily obtain extremely large multiples of compression. Therefore, these methods can often only prove that a certain method is correct, but it is of little significance. The more meaningful challenge is to compress those models that are inherently more efficient in terms of speed and accuracy trade-offs. This article tries to simplify the Mobile Net itself. The channel pruning operation is performed on the network. The main structure in Mobile Net is a depth separable convolution unit, so this paper proposes a channel pruning method based on a depth separable convolution unit.

(1) Pruning process
The depth separable convolution unit is composed of multiple 3 × 3 spatial convolution layers and multiple 1 × 1 channel convolution layers. The main calculation amount of this unit is concentrated in the 1 × 1 channel convolution. If the pruning operation is performed on the channels obtained by the 3 × 3 spatial convolution layer, the number of input channels of the 1 × 1

convolution layer can be significantly reduced, thereby reducing the computational complexity of the 1×1 channel convolution.

Now we consider explaining this method from another angle. In the depth separable convolution, the 3×3 spatial convolution can be regarded as the feature extraction process, and the features of each channel of the feature layer are filtered to obtain the features. After the features are obtained, the features are performed by the 1×1 channel convolution method. The introduction of channel pruning after the spatial convolutional layer can be regarded as introducing a feature selection process to retain the more important features in image quantization, detection or segmentation tasks, and filter out the less important features.

We use a triple $<$ Li,Di,Pi $>$ to represent the i-th depth separable unit, Li refers to the input unit, Di refers to the spatial convolution in the depth separable convolution structure, and Pi refers to the channel convolution in the structure, namely convolution part. The activation layer in the figure refers to the activation layer after the Di layer, which is the input unit of channel convolution. The goal of channel pruning is to cut the less important channels in the activation layer. At the same time, the corresponding convolutional layers in Di and Pi will also be removed.

Each filter in Di corresponds to a channel in the activation layer, and the importance of the feature channel is often evaluated by some index. The light-colored layer in the activation layer represents the output channel with lower importance and should be removed. The corresponding convolution filter in Di is removed, and the convolution filter that uses this channel as input is removed at the same time, and the output dimension of the depth separable convolution unit remains unchanged. From the calculation principle of the depth separable convolution, we know that a certain channel j of the activation layer is obtained by convolution of the j-th channel of the input unit with the 3×3 filter of the j-channel of the spatial convolution layer. The channel is removed, which means that the j-th channel of the input layer should also be removed. For the convolution filter that should be removed, in the pruning process, the corresponding convolutional layer weight value is set to 0, and the corresponding learning rate is also set to 0. Therefore, the training and inference process of the convolutional neural network is not in calculations. After the network fine-tuning training is completed, the convolutional filter with an ownership value of 0 in the network will be removed. At this point, the unimportant convolutional layer in the neural network has been completely removed. The calculation amount of the entire module has been calculated as $D1 \times D1 \times D2 \times D2 \times M + M \times N \times D1 \times D1$. Now some channels are removed by the method of channel pruning. It should be noted that the pruning process does not reduce the output dimension of the convolution unit, but only cuts off the less important feature layer and its corresponding weight. Therefore, assuming that $M \times \varepsilon$ channels ($0 < \varepsilon < 1$) are removed, the entire depth can be separated, the calculation amount of the convolution unit becomes $D1 \times D1 \times D2 \times D2 \times M \times (1 - \varepsilon) + M \times N \times D1 \times D1 \times (1 - \varepsilon)$, which is greatly reduced compared to the original convolutional neural network. And the output dimension size N will not change.

The pruning of the i-th depth separable convolution unit will also affect the i-1th convolution unit. Assume that the j channel of the input layer Li of the i-th depth separable convolution unit is removed, and Li is generated by the channel convolution part Pi-1 of the i-1th depth separable convolution structure, so the corresponding 1×1 convolution part can also be removed. It can be seen that the channel pruning method based on the depth separable unit has better interpretability than the ordinary convolution unit, especially when dealing with multilayer structure pruning.

A complete channel pruning process based on depth separable convolution unit is shown in **Figure 5**. For the i-th depth separable convolution unit $<$ Li,Di,Pi $>$, all channels of the spatial convolution layer Di output are obtained according to a certain importance evaluation criterion degree of importance. Then you sort the convolution channels in the order of importance from small to large, remove some of the least important channels in the output, and then remove the corresponding convolution filters in Pi and Di and the corresponding ones in Li input channel. You remove the convolution filter corresponding to Li in the layer $<$ Li-1,Di-1,Pi-1 $>$ of the depth separable convolution unit, so that the pruning of the i-th depth separable convolution unit process is over. After the pruning of all convolutional units is completed, the network fine-tuning method is used to restore the lost performance of the model after pruning.

(2) Evaluation of channel importance and selection of pruning ratio

The first and most important step of channel pruning is to evaluate the importance of feature channels. Important feature channels retain some of the more important features in the model. If removed, the accuracy of the model will be greatly affected. This article chooses APo Z as the evaluation standard for the importance of the convolution channel. In the depth separable convolution unit, the spatial 3×3 convolution layer is convolved separately for each input channel to obtain image features. Therefore, there is a one-to-one correspondence between the convolution channel and the output feature layer. After the feature of the spatial convolutional layer Di is taken, nonlinearity is introduced through the linear rectification activation function (Re Lu) to obtain the activation layer Ai. Therefore, the higher the APo Z in the activation layer, the greater the proportion of 0 in the activation layer, indicating the most positions of the feature layer are not activated, and their importance is low, and their corresponding spatial convolutional layer Di should be removed.

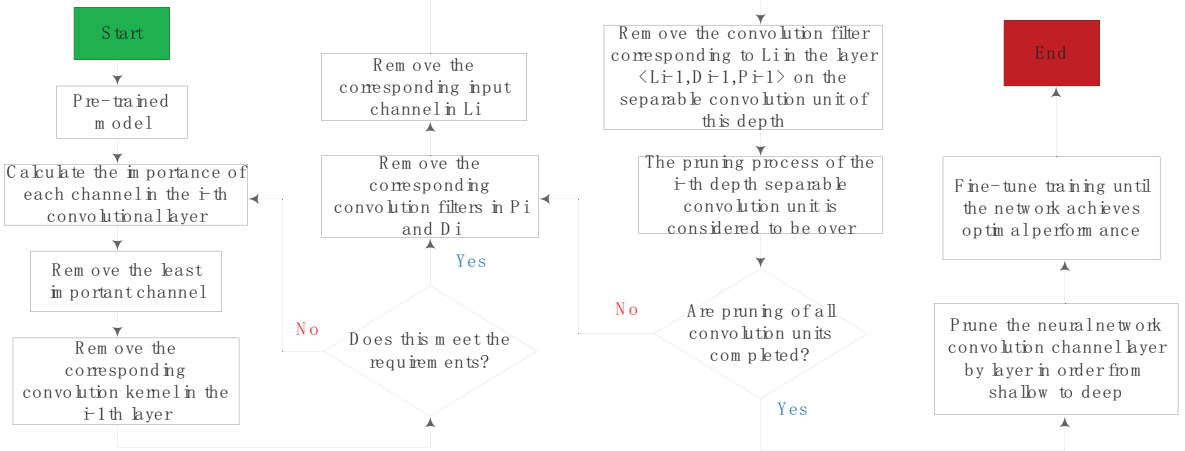


FIGURE 5 | The overall process of channel pruning.

The Softmax function, also known as the normalized exponential function, is a generalization of the logical function, and the function form is as follows:

$$\sigma(z)_j = \frac{e^{-zj}}{\sum_{k=1}^{K} e^{-zk}} \tag{7}$$

It can be seen from the formula that the Softmax function is actually the normalization of the discrete probability distribution of finite items, and Softmax is widely used in neural networks for multi-quantization problems.

It can be seen from the Softmax formula that the Softmax function is actually the gradient log normalization of the discrete probability distribution of finite items. Softmax is widely used in neural networks for multi-quantization problems. According to the Softmax function, the probability value corresponding to each soft label is obtained. In deep learning tasks, the goal of training is to optimize the loss function, and maximum likelihood estimation is generally used to construct the loss function in quantization problems. For the input x, assuming our class label is t, the goal of the quantization task is to find the most suitable model to maximize p(t| x). In the second quantification problem, we can know from the probability knowledge:

$$p\,(t\,|x) = (1-y)^{t-1} \cdot y^{t} \tag{8}$$

y = f(x) is the probability value predicted by the model, and t is the class label corresponding to the sample. Generalizing the two-quantization problem to a more general multi-quantization problem, p(t| x) is expressed as follows:

$$p\,(t\,|x) = \prod_{i=0}^{C-1} p\,(t_i\,|x)^{t_i-1} = \prod_{i=0}^{C-1} y_i^{t_i-1} \tag{9}$$

In actual calculations, continuous multiplication may cause the final result to approach 0, and in the process of network back propagation, the exponential function is inconvenient to handle, so the likelihood function is generally taken as the negative logarithm of the logarithmic likelihood function to convert the problem of maximizing p(t| x) into a problem of minimizing the log-likelihood function.

$$L(t,x) = -t \cdot \log p(t\,|x) = -\sum_{i=0}^{C-1} t_i \cdot \log y_i \tag{10}$$

Using the cross entropy function not only can measure the effect of the model very well, but also can easily calculate the derivative.

Now we need to select a data set, and get the output of each channel after activation of the spatial convolution layer Di for each image sample of the data set. We calculate the number of activation values of 0 in the feature map, divide it by the size of the feature map, and finally average all samples to get the APo Z value of the convolution channel corresponding to the spatial convolution layer. After that, we sort the feature channels in the order of APo Z from small to large, and remove the M × ε channel with the largest APo Z value according to a certain pruning ratio ε, and retain the convolution channel with the smaller APo Z value. The APo Z-based convolutional channel importance evaluation algorithm has the advantages of simple method and fast calculation, and is widely used in convolutional neural network channel pruning tasks.

According to APo Z, channels with low APo Z retain more effective information than channels with high APo Z. A spatial layer will output hundreds or thousands of characteristic channels. The more channels removed, the more effective the network performance, the greater the impact. There is currently no standard answer to this question. Different models and different data sets will have different pruning ratios.

The pruning of the upper layer of convolutional unit will have a certain impact on the next layer of convolutional layer unit, so there are two methods for the performance recovery after model pruning, one is the greedy pruning method, that is, the upper and lower layers are not considered as a result, all convolutional channels of the network are pruned at one time, and then the network is fine-tuned to restore network accuracy. Another way

is to prun and fine-tune the neural network convolution channel layer by layer from shallow to deep until the network achieves the best performance. Both methods have their own advantages and disadvantages. The process of the greedy pruning algorithm is relatively simple, and the pruning steps and fine-tuning steps are time-consuming; the advantage of the layer-by-layer pruning method is that it is often more accurate.

COGNITIVE FUNCTION QUANTITATIVE EXPERIMENT ANALYSIS

Experimental Data

All the data in the experiment comes from the ADNI database, and the age information of the subjects is shown in **Figures 6**, **7**. We use all the initial moment data. There are a total of 800 subject samples containing MRI data, including 168 AD, 403 MCI and 229 NC, and a total of 339 multimodal data samples containing both MRI and PET, including 93 AD, 146 MCI and 100 NC.

The original image size is 256 × 256 × 256, and the surrounding pixels without information are removed after sampling by down 2. The maximum enclosing size of MRI and PET is 100 × 81 × 80. The cross-validation method is used to do ten-fold cross-validation. The training data set is enhanced by shifting sampling in all directions, which is increased by eight times. The verification data set and the test data set are not enhanced. We use the Adadelta gradient optimization algorithm to learn the weights of the network, and the batch size is set to 64. During training, the model tends to converge about 20 iterations. In this experiment, we will take several experiments to evaluate the model comprehensively on the three quantitative tasks of AD vs. NC, p MCI vs. NC and s MCI vs. NC. When training local 3D-CNN, the network weight of AD vs. NC is initialized in the same way as 2D-CNN, using Xavier for random initialization. Since the difference between mild cognitive impairment samples and NC is small, the transfer learning method is used to alleviate the problem of insufficient training caused by the small amount of training data. We will use the trained AD vs. NC network for the p MCI vs. NC network. The NC network is initialized. In the same way, the network of s MCI vs. NC is initialized with the network of p MCI vs. NC.

In order to more comprehensively verify the effectiveness of the cascaded neural network proposed in this paper, we launched a comparative experiment. It is worth noting that the following MRI monomodal experimental results are based on all the 800 MRI images at the initial moments. The PET monomodal and multimodal experiments used 339 recipients who participated in both MRI and PET scans. The following are the results and analysis of each experiment.

Parallel 3D-CNNs Experiment

We use single-modal images to quantify AD vs. NC, and compare the results of benchmark tests, single CNN quantification of the entire brain image, and single CNN quantification of the performance of partial images. **Figure 8** shows the comparison results. Among them, the method used in "benchmark test" is to calculate Propensity Score Matching (p-score) based on two categories, obtain the probability estimation of each pixel, and use the verification set to select the first n pixels with the most discriminative ability, and then use the support vector machine to quantify the result. "A single image block with the best performance" represents the block with the highest quantization accuracy among the 27 image blocks. "Single CNN quantization" downsamples the entire brain image by 4 and inputs it to the same network as the quantized partial image block for quantization. "Parallel 3D-CNNs integration method" refers to the network structure.

From the experimental results in the observation table, it can be found that the 3D convolutional neural network model proposed in this article has a significant improvement in the effect of the benchmark test. The method of integrating 3D-CNNs can further improve the prediction of AD disease in a single modal 3D image. These evaluation indicators verify the effectiveness of the parallel multiple CNNs model proposed in this paper.

Multimodal Fusion Experiment

We combined the local features of MRI and PET, and used the multi-modal data fusion method of cascaded

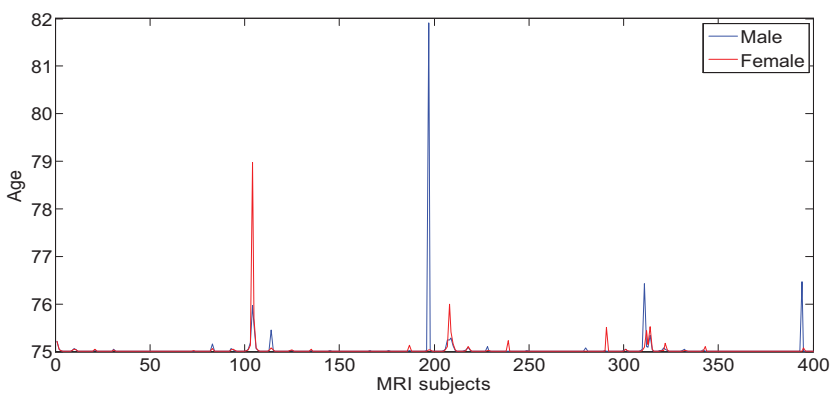


FIGURE 6 | Age characteristics of Alzheimer's disease of MRI subjects.

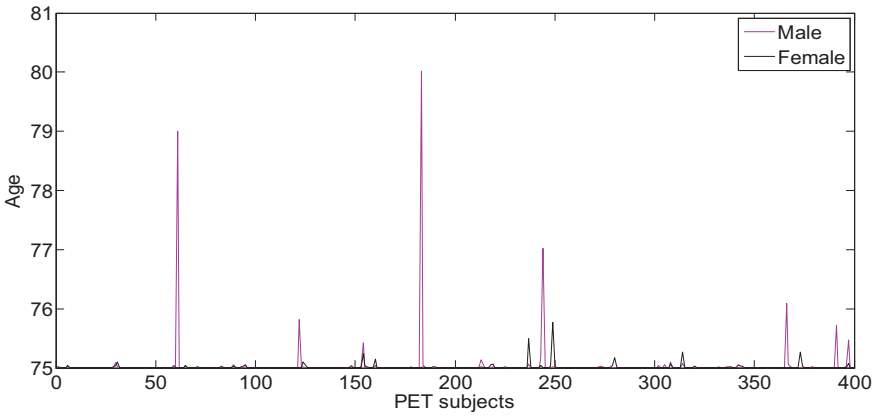


FIGURE 7 | Age characteristics of Alzheimer's disease of PET subjects.

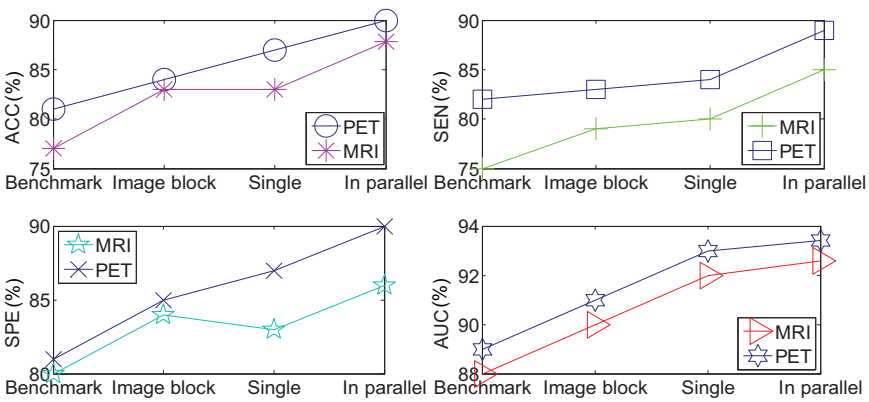


FIGURE 8 | Comparison of parallel 3D-CNNs integration method and other methods.

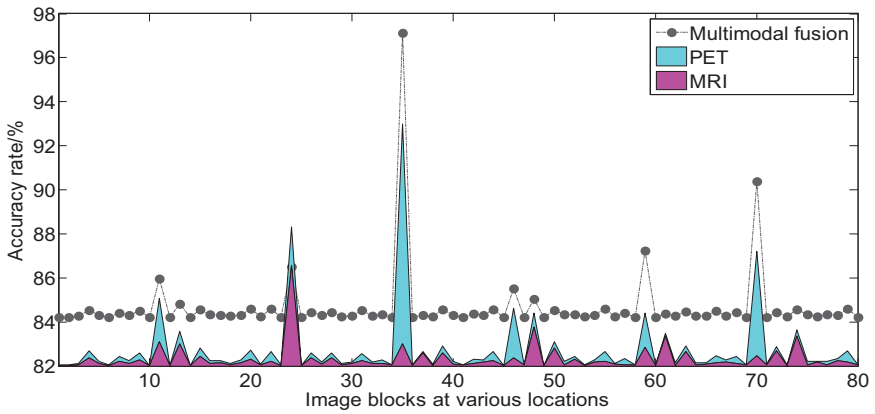


FIGURE 9 | Comparison of AD vs. NC accuracy of 3D image blocks at various positions before and after multi-modal fusion.

multi-CNNs to quantify AD vs. NC, p MCI vs. NC and s MCI vs. NC, respectively, to verify the advantages of single-mode analysis.

We take AD vs. NC as an example to analyze the multi-modal fusion effect of a single location image block. The histogram shown in **Figure 9** represents the accuracy results of the multi-modal fusion of 80 image blocks.

From the histogram shown in **Figure 9**, we can find that we use 2D-CNN to perform feature fusion of the local image blocks of the two modalities compared to the accuracy of

the quantization of a single mode, which can improve the quantization performance in almost all positions. It can be seen that the multimodal fusion method proposed in this paper is effective.

We then quantify the features of multiple local image blocks through a fully connected layer combination. **Figure 10** shows the three quantization tasks of AD vs. NC, p MCI vs. NC and s MCI vs. NC in single mode and multi-mode. Among them, the results of "MRI" and "PET" in the figure are obtained directly from a single mode image. It can be seen from the results in **Figure 10** that the quantified performance of the multi-modal data is greatly improved in all indicators compared with the single-modal data.

We drew the ROC curves of AD vs. NC, p MCI vs. NC, and s MCI vs. NC based on the network prediction results, as shown in **Figures 11–13**, respectively. Observing the ROC curve, it can be found that the performance of multimodal fusion is significantly better than that of single mode, and the performance of PET is slightly lower than the quantitative performance of MRI.

Comparison With Other Fusion Algorithms

The features extracted based on the network are compared with the multi-modal fusion method with multiple feature fusion methods such as direct averaging method, feature stitching method, parallel method, and feature bilinear point multiplication. As shown in **Figure 14**, "average" refers to the average predicted probability of a single sample using two modalities as the final predicted probability of the sample, and "splicing" means concatenating the features of the fully connected layers of the two modalities into one. The dimensional features are re-quantified as the total features of the sample. "Parallel" means to connect the features of the last convolutional layer in

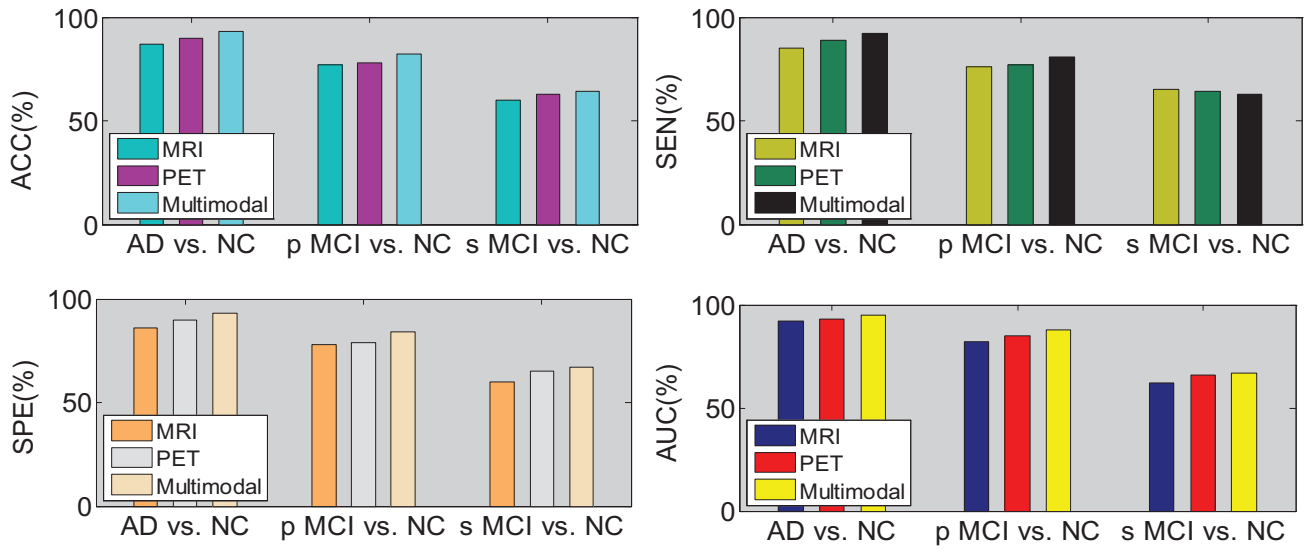


FIGURE 10 | Performance comparison of single-mode and multi-mode under three quantitative tasks.

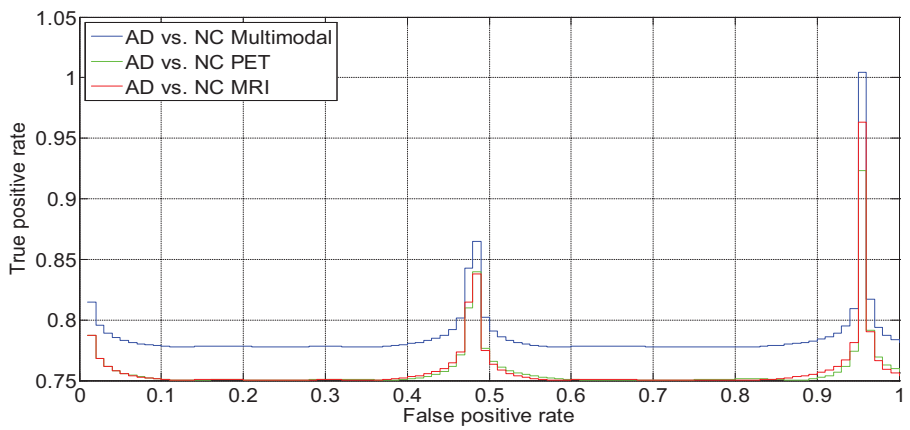


FIGURE 11 | AD vs. NC fusion quantified ROC curve.

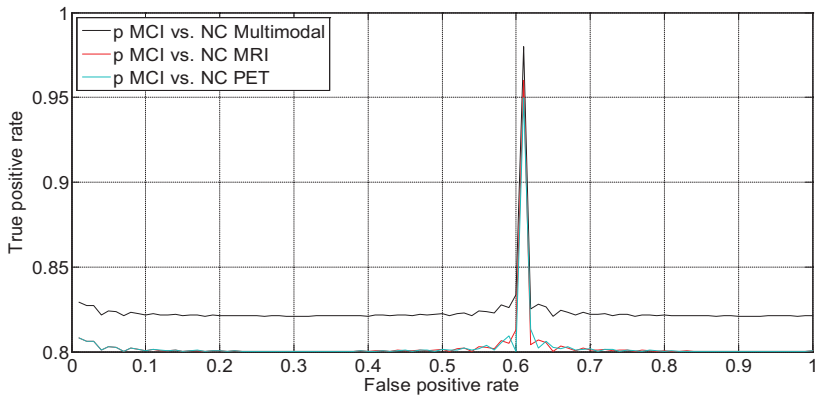


FIGURE 12 | ROC curve of p MCI vs. NC fusion quantification.

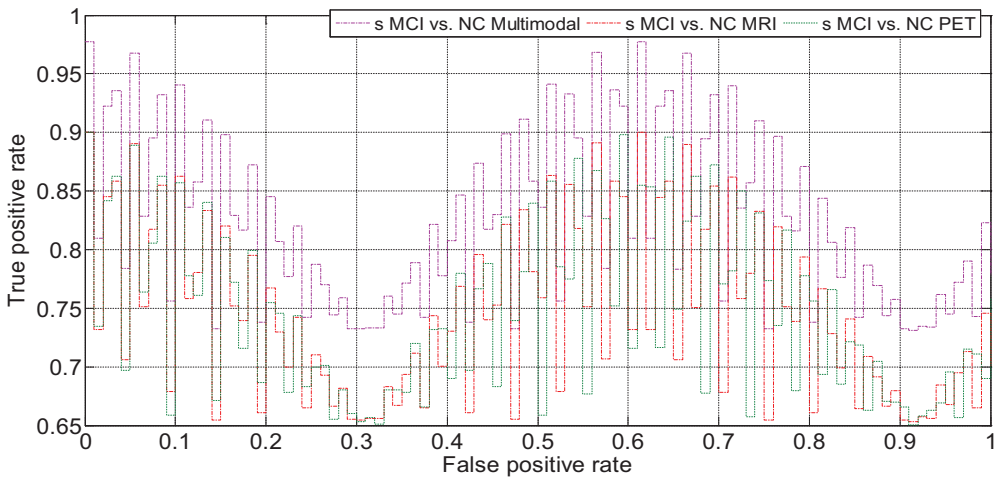


FIGURE 13 | s MCI vs. NC fusion quantified ROC curve.

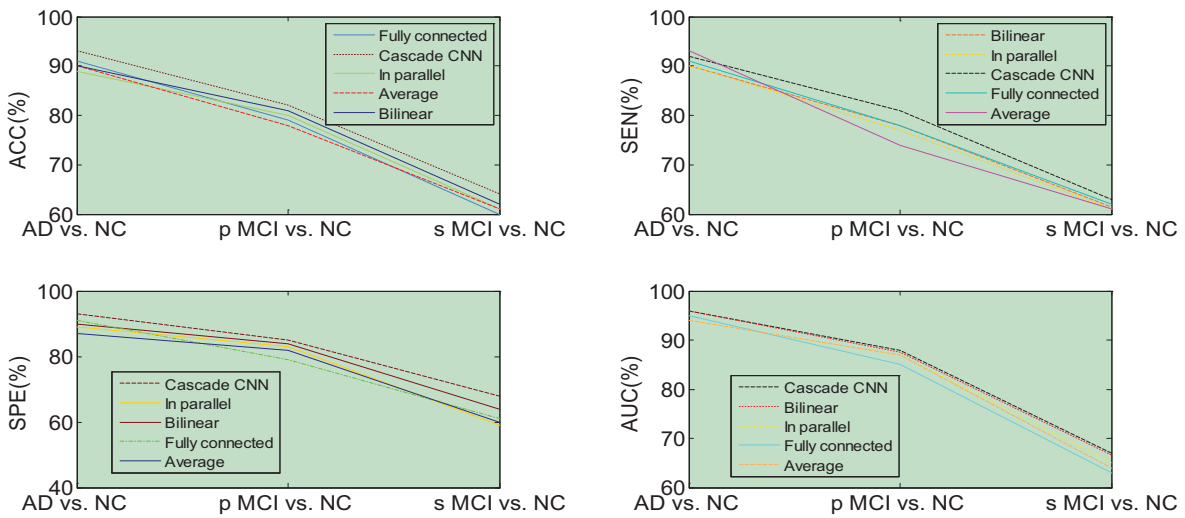


FIGURE 14 | Performance comparison of five multi-modal fusion methods.

parallel and then perform subsequent quantification. "Bilinear" means to use the features of the fully connected layers of MRI and PET for bilinear multiplication, which is fused into a feature matrix for processing.

The experimental results shown in **Figure 14** show that the performance of our proposed multimodal fusion method is better than other methods for each quantitative task. The above results show that our proposed multi-modal fusion method can capture the correlation of high-level features.

Discriminative Quantitative Analysis

Different from traditional manual feature extraction methods, the features extracted in this paper are automatically learned in image quantization through cascaded neural networks. The gray information on the original image is transformed from shallow to high-level by non-linear transformation of the cascaded neural network, gradually transformed into high-level information with discriminative ability, and finally these expressive features are mapped to disease prediction. In the neural network, the direct display of these "high-level" hidden layer features is usually poor in interpretability. In the quantification of medical images, the neuroimaging is calculated to qualitatively and quantitatively analyze the relevant brain regions affected by AD disease. It is very important to analyze the pathological causes of Alzheimer's disease and drug research.

Aiming at the problem of poor feature interpretation of deep convolutional neural networks, we try to visualize the areas that have a large impact on quantitative prediction. In order to achieve this goal, we systematically slide the 3D gray cubes on the original image to cover the information of different brain regions, and

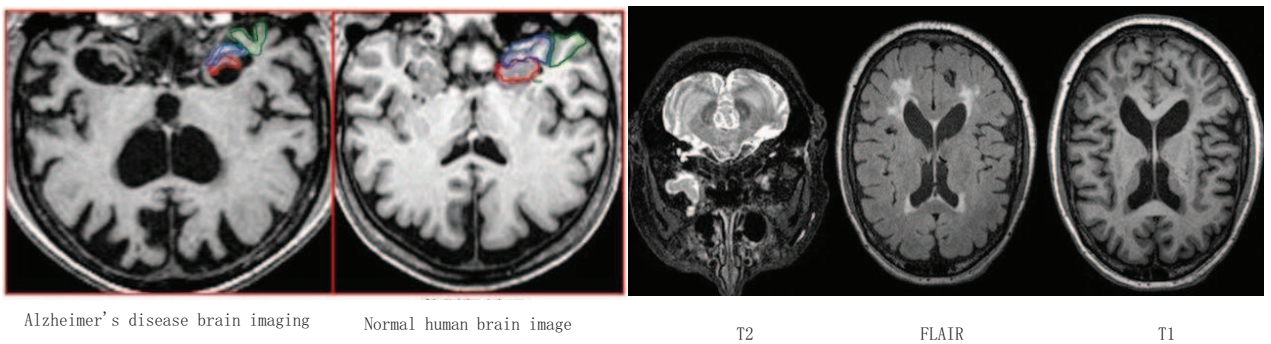


FIGURE 15 | Alzheimer's disease area that the neural network focuses on.

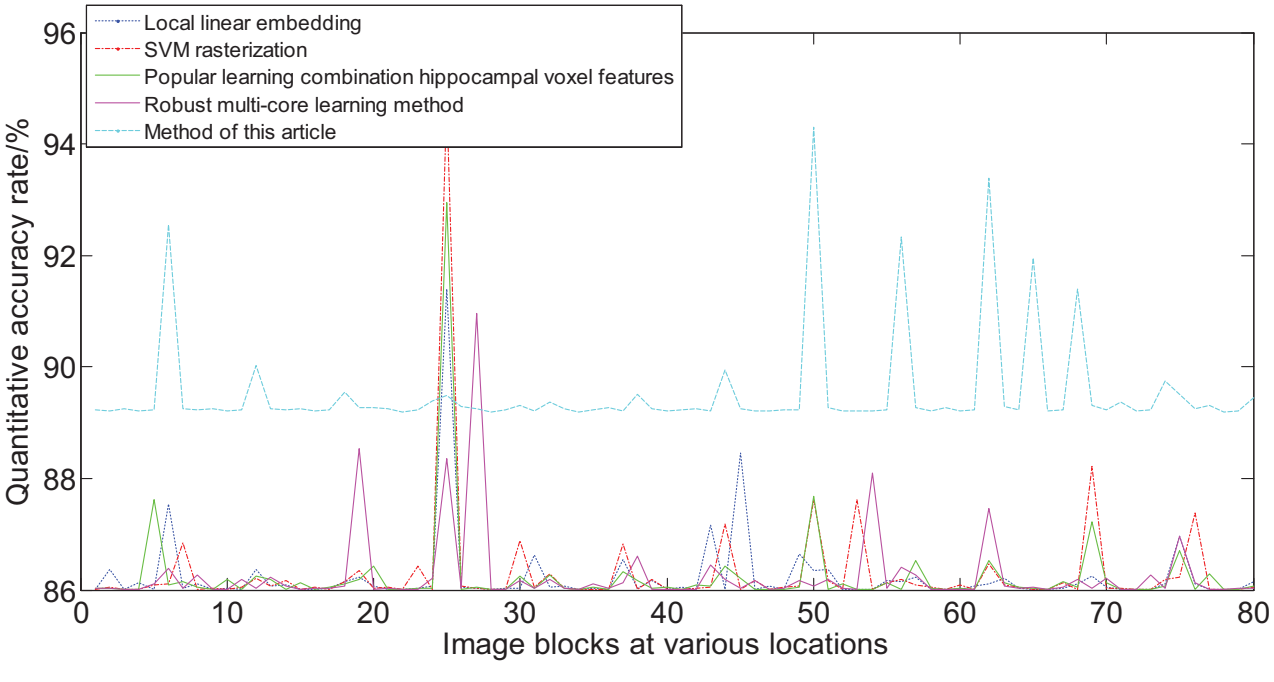


FIGURE 16 | Comparison of the quantification accuracy of the multi-modal cascaded 3D-CNNs proposed in this article and other methods.

use the trained model to monitor the changes in the probability output of the network quantification under different masking positions of the original image. If the excluded patch covers the area related to AD, the predicted probability of the correct class will decrease significantly. Therefore, by predicting the change in probability, we can roughly determine the key focus areas that have an impact on AD.

In the experiment, we used MRI and PET images of 10 Alzheimer's disease patient samples in the test set for visual analysis. First, we select three image blocks with the best overall quantization performance in the two modalities for analysis. Then, each image block uses a $15 \times 15 \times 15$ gray cube block (the gray value is the mean value of the input image block) to slide and mask each area on the image block. Probability prediction is made in the corresponding 3D neural network model. Finally, we calculate the decrease in AD prediction probability after masking to generate the probability change of each sample in each region.

Figure 15 shows the area that the neural network focuses on the MRI image. The red part represents the area with the largest probability change, that is, the structural position most relevant to AD. Furthermore, we compare the experimental results of this method with the results of other methods. **Figure 16** shows the overall accuracy of the quantification results, comparing the overall results of multimodal quantification for AD vs. NC, p MCI vs. NC, and s MCI vs. NC. It can be seen from the results in the figure that the method proposed in this paper is better than other methods in the quantification of AD vs. NC, and has better sensitivity and greater ROC curve line in MCI vs. NC. The area helps to detect MCI earlier and prevent missed diagnosis. These results once again prove the effectiveness of the method proposed in this paper.

CONCLUSION

In this paper, DTI, f MRI, and T1 data are preprocessed and the brain function connection network and the brain white matter structure connection network are respectively constructed. Then the graph theory parameters are introduced into the analysis of the two kinds of brain networks, and the graph theory parameters with significant differences between AD patients and normal people are found as features by T test. This article introduces the calculation principle of the efficient convolution structure-depth separable convolution unit existing in Mobile Net, analyzes its computational complexity and compares it with the traditional convolution layer, and explains that the depth separable convolution unit is in the convolution efficiency in product operations. On this basis, a channel pruning method based on a depth separable convolution unit is presented. The pruning process and the effect of network compression are analyzed with the help of a flowchart. This article is the use of single, parallel and cascaded convolutional neural networks to extract and quantify the experimental process and results of FDG-PET images and MRI. The first is the introduction to the experiment, which describes the sample selection we used for the experiment and the preprocessing work done before the sample is input to the network. The comparative experiment of the evaluation model is introduced, and the experimental results are given to verify the effectiveness of our algorithm framework. This article analyzes the experimental results, and uses the trained model to reversely analyze the affected area of CNN to infer the discriminative brain areas of Alzheimer's disease. The results prove that the method of combining 2D-CNNs and BGRU proposed in this paper has excellent distinguishing ability for PET monomodal data, indicating that the combination of CNN and RNN can capture the functional change information of brain images. The model of cascaded three-dimensional convolutional neural network can fully utilize the information of multi-modal data while extracting single-modal structural features. The brain image calculation and analysis model proposed in this paper can accurately and effectively quantify the cognitive function of Alzheimer's disease according to the characteristics of the modal.

DATA AVAILABILITY STATEMENT

The dataset analyzed for this study can be found in this link. http://adni.loni.usc.edu/.

AUTHOR CONTRIBUTIONS

YH and JW contributed equally to this work. All authors contributed to the article and approved the submitted version.

REFERENCES

Bi, X., Li, S., Xiao, B., Li, Y., Wang, G., and Ma, X. (2020a). Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. *Neurocomputing* 392, 296–304. doi: 10.1016/j.neucom.2018.11.111

Bi, X., Zhao, X., Huang, H., Chen, D., and Ma, Y. (2020b). Functional brain network classification for alzheimer's disease detection with deep features and extreme learning machine. *Cogn. Comput.* 12, 513–527. doi: 10.1007/s12559-019-09688-2

Chen, H. Y., Chen, J. Q., Li, J. Y., Huang, H. J., Chen, X., Zhang, H. Y., et al. (2019). Deep learning and random forest approach for finding the optimal traditional chinese medicine formula for treatment of alzheimer's disease. *J. Chem. Inform. Modeling* 59, 1605–1623. doi: 10.1021/acs.jcim.9b00041

Choi, H., Ha, S., Kang, H., Lee, H., Lee, D. S., Alzheimer's Disease et al. (2019). Deep learning only by normal brain PET identify unheralded brain anomalies. *EBioMedicine* 43, 447–453. doi: 10.1016/j.ebiom.2019.04.022

Choi, H., Jin, K. H., and Alzheimer's Disease Neuroimaging Initiative (2018). Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav. Brain Res.* 344, 103–109. doi: 10.1016/j.bbr.2018.02.017

El-Sappagh, S., Abuhmed, T., Islam, S. R., and Kwak, K. S. (2020). Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. *Neurocomputing* 412, 197–215. doi: 10.1016/j.neucom.2020.05.087

Gulhare, K. K., Shukla, S. P., and Sharma, L. K. (2017). Deep neural network classification method to Alzheimer's disease detection. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 7, 1–4. doi: 10.23956/ijarcsse/v7i6/0259

Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit. Med.* 3, 1–9. doi: 10.5121/hiij.2014.3201

Jain, R., Jain, N., Aggarwal, A., and Hemanth, D. J. (2019). Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. *Cogn. Syst. Res.* 57, 147–159. doi: 10.1016/j.cogsys.2018.12.015

Jha, D., and Kwon, G. (2017). Alzheimer's disease detection using sparse autoencoder, scale conjugate gradient and softmax output layer with fine tuning. *Int. J. Mach. Learn. Comput.* 7, 13–17. doi: 10.18178/ijmlc.2017.7.1.612

Ju, R., Hu, C., and Li, Q. (2017). Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16, 244–257. doi: 10.1109/tcbb.2017.2776910

Kam, T. E., Zhang, H., Jiao, Z., and Shen, D. (2019). Deep learning of static and dynamic brain functional networks for early mci detection. *IEEE Trans. Med. Imaging* 39, 478–487. doi: 10.1109/tmi.2019.2928790

Kumar, P. J., Yung, Y., and Huan, T. L. (2017). Neural network based decision trees using machine learning for alzheimer's diagnosis. *Int. J. Comput. Inform. Sci.* 4, 63–72.

Le, N. Q. K., Huynh, T. T., Yapp, E. K. Y., and Yeh, H. Y. (2019). Identification of clathrin proteins by incorporating hyperparameter optimization in deep learning and PSSM profiles. *Comput. Methods Programs* 177, 81–88. doi: 10.1016/j.cmpb.2019.05.016

Lee, G., Nho, K., Kang, B., Sohn, K. A., and Kim, D. (2019). Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Sci. Rep.* 9:1952.

Li, F., Liu, M., Alzheimer's Disease Neuroimaging Initiative (2018). Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. *Comput. Med. Imaging Graph.* 70, 101–110. doi: 10.1016/j.compmedimag.2018.09.009

Li, H., Habes, M., Wolk, D. A., Fan, Y., and Alzheimer's Disease Neuroimaging Initiative (2019). A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's Dement.* 15, 1059–1070. doi: 10.1016/j.jalz.2019.02.007

Liu, M., Zhang, J., Adeli, E., and Shen, D. (2018). Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. *IEEE Trans. Biomed. Eng.* 66, 1195–1206. doi: 10.1109/tbme.2018.2869989

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018a). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci. Rep.* 8:5697.

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018b). Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. *Med. Image Anal.* 46, 26–34. doi: 10.1016/j.media.2018.02.002

Luo, S., Li, X., and Li, J. (2017). Automatic Alzheimer's disease recognition from MRI data using deep learning method. *J. Appl. Math. Phys.* 5, 1892–1898. doi: 10.4236/jamp.2017.59159

Martinez-Murcia, F. J., Ortiz, A., Gorriz, J. M., Ramirez, J., and Castillo-Barnes, D. (2019). Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders. *IEEE J. Biomed. Health Inform.* 24, 17–26. doi: 10.1109/jbhi.2019.2914970

Mazurowski, M. A., Buda, M., Saha, A., and Bashir, M. R. (2019). Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. *J. Magn. Reson. Imaging* 49, 939–954. doi: 10.1002/jmri.26534

Noor, M. B. T., Zenia, N. Z., Kaiser, M. S., Al Mamun, S., and Mahmud, M. (2020). Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Inform.* 7:11.

Oh, K., Chung, Y. C., Kim, K. W., Kim, W. S., and Oh, I. S. (2019). Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. *Sci. Rep.* 9:18150.

Raza, M., Awais, M., Ellahi, W., Aslam, N., Nguyen, H. X., and Le-Minh, H. (2019). Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques. *Expert Syst. Appl.* 136, 353–364. doi: 10.1016/j.eswa.2019.06.038

Shi, J., Zheng, X., Li, Y., Zhang, Q., and Ying, S. (2017). Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE J. Biomed. Health Inform.* 22, 173–183. doi: 10.1109/jbhi.2017.2655720

Spasov, S., Passamonti, L., Duggento, A., Liò, P., Toschi, N., and Alzheimer's Disease Neuroimaging Initiative (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage* 189, 276–287. doi: 10.1016/j.neuroimage.2019.01.031

Tang, Z., Chuang, K. V., DeCarli, C., Jin, L. W., Beckett, L., Keiser, M. J., et al. (2019). Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline. *Nat. Commun.* 10:2173.

Vieira, S., Pinaya, W. H., and Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. *Neurosci. Biobehav. Rev.* 74, 58–75. doi: 10.1016/j.neubiorev.2017.01.002

Vu, T. D., Ho, N. H., Yang, H. J., Kim, J., and Song, H. C. (2018). Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection. *Soft Comput.* 22, 6825–6833. doi: 10.1007/s00500-018-3421-5

Wang, H., Shen, Y., Wang, S., Xiao, T., Deng, L., Wang, X., et al. (2019). Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. *Neurocomputing* 333, 145–156. doi: 10.1016/j.neucom.2018.12.018

Wingate, J., Kollia, I., Bidaut, L., and Kollias, S. (2020). Unified deep learning approach for prediction of Parkinson's disease. *IET Image Process.* 14, 1980–1989. doi: 10.1049/iet-ipr.2019.1526

Zeng, N., Qiu, H., Wang, Z., Liu, W., Zhang, H., and Li, Y. (2018). A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing* 320, 195–202. doi: 10.1016/j.neucom.2018.09.001

Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., and Cheng, F. (2019). deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics* 35, 5191–5198. doi: 10.1093/bioinformatics/btz418

Zhang, Q., Bai, C., Chen, Z., Li, P., Wang, S., and Gao, H. (2019). Smart Chinese medicine for hypertension treatment with a deep learning model. *J. Netw. Comput. Appl.* 129, 1–8. doi: 10.1016/j.jnca.2018.12.012

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 March 2021
doi: 10.3389/fnins.2021.651574

A Novel Unit-Based Personalized Fingerprint Feature Selection Strategy for Dynamic Functional Connectivity Networks

Feng Zhao[1], Zhiyuan Chen[1], Islem Rekik[2,3], Peiqiang Liu[1], Ning Mao[4], Seong-Whan Lee[7] and Dinggang Shen[5,6,7]**

[1] School of Computer Science and Technology, Shandong Technology and Business University, Yantai, China, [2] BASIRA Lab, Faculty of Computer and Informatics Engineering, Istanbul Technical University, Istanbul, Turkey, [3] School of Science and Engineering, Computing, University of Dundee, Dundee, United Kingdom, [4] Department of Radiology, Yantai Yuhuangding Hospital, Yantai, China, [5] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China, [6] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China, [7] Department of Artificial Intelligence, Korea University, Seoul, South Korea

The sliding-window-based dynamic functional connectivity networks (SW-D-FCN) derive from resting-state functional Magnetic Resonance Imaging has become an increasingly useful tool in the diagnosis of various neurodegenerative diseases. However, it is still challenging to learn how to extract and select the most discriminative features from SW-D-FCN. Conventionally, existing methods opt to select a single discriminative feature set or concatenate a few more from the SW-D-FCN. However, such reductionist strategies may fail to fully capture the personalized discriminative characteristics contained in each functional connectivity (FC) sequence of the SW-D-FCN. To address this issue, we propose a unit-based personalized fingerprint feature selection (UPFFS) strategy to better capture the most discriminative feature associated with a target disease for each unit. Specifically, we regard the FC sequence between any pair of brain regions of interest (ROIs) is regarded as a unit. For each unit, the most discriminative feature is identified by a specific feature evaluation method and all the most discriminative features are then concatenated together as a feature set for the subsequent classification task. In such a way, the personalized fingerprint feature derived from each FC sequence can be fully mined and utilized in classification decision. To illustrate the effectiveness of the proposed strategy, we conduct experiments to distinguish subjects diagnosed with autism spectrum disorder from normal controls. Experimental results show that the proposed strategy can select relevant discriminative features and achieve superior performance to benchmark methods.

Keywords: dynamic functional connectivity networks, resting-state functional magnetic resonance imaging, feature selection strategy, functional connectivity network, autism spectrum disorder

OPEN ACCESS

Edited by:
Yizhang Jiang, Jiangnan University, China

Reviewed by:
Qiang Guo, Shandong University of Finance and Economics, China
Qiguang Miao, Xidian University, China

****Correspondence:***
Feng Zhao
zhaofeng1016@126.com
Dinggang Shen
dinggang.shen@gmail.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *10 January 2021*
Accepted: *15 February 2021*
Published: *22 March 2021*

Citation:
Zhao F, Chen Z, Rekik I, Liu P, Mao N, Lee S-W and Shen D (2021) A Novel Unit-Based Personalized Fingerprint Feature Selection Strategy for Dynamic Functional Connectivity Networks. Front. Neurosci. 15:651574. doi: 10.3389/fnins.2021.651574

INTRODUCTION

The resting-state functional Magnetic Resonance Imaging (rs-fMRI), as a non-invasive neuroimaging technique, has been widely applied to capture the blood-oxygen-level-dependent (BOLD) signal which is sensitive to the spontaneous and intrinsic neural activity within the brain (Liu et al., 2008). In particular, the functional connectivity (FC), which is defined as the temporal

correlation of the rs-fMRI signals between different brain regions of interest [ROIs, which is often defined through image registration (Luan et al., 2008; Jia et al., 2010; Wu et al., 2011, 2013)], can reflect the degree to which ROIs co-interact (Van Den Heuvel and Hulshoff Pol, 2010; Biao et al., 2016). A functional connectivity network (FCN) is usually represented as a graph, where each node represents a brain ROI and the edge between two nodes encodes the associated FC. Nowadays, many FCN models have been developed, from the simple conventional FCN (Zhang et al., 2015, 2016; Qiao et al., 2018) to the complex time-frequency analysis based dynamic FCN (Yaesoubi et al., 2015; Du et al., 2018), for diagnosing some neurodevelopmental diseases including the autism spectrum disorder (ASD) (Fornito et al., 2015; Liu et al., 2016; Huang et al., 2018; Zhao et al., 2018), the major depressive disorder (Greicius et al., 2007; Cullen et al., 2014) and so on.

Recent studies have shown that the dynamic changes of the correlation between ROIs contain abundant information (Damaraju et al., 2014; Wee et al., 2015). To explore the relationship between the dynamic changes and brain diseases while accounting for the time-varying connectivity patterns across ROIs, the sliding window based D-FCN (SW-D-FCN) is the most widely used technique on account of its simplicity and effectiveness (Chen et al., 2016, 2017; Du et al., 2018; Zhou et al., 2018). Specifically the sliding window approach is firstly used to generate a set of rs-fMRI subseries. Next, Pearson's Correlation (PC) based temporal FCN is generated for each subseries. In the SW-D-FCN, the temporal change of the correlation between a pair of ROIs along the scanning time is defined as an FC sequence and thus SW-D-FCN also can be regarded as a set of FC sequences.

Feature extraction and selection are important in the process of exploring the time-varying connectivity patterns implicated in SW-D-FCN. Recently, many different feature extraction methods have been proposed. For example, Zhang et al. computed a weighted local clustering coefficient for each node in each network as a feature (Zhang et al., 2017); Chen et al. extracted features by calculating the root-mean-square for each FC sequence (Chen et al., 2017). Although many features can be obtained by various feature extraction methods to reflect the dynamic changes of an FC sequence, those features are prone to redundancy, which hinders the learning of robust classifiers. More importantly, many features may be irrelevant to the diagnosis of specific disease. Therefore, to improve the diagnosis performance, what kind of feature selection strategy should be adopted to select a smaller subset from the big feature set remains a key issue to solve.

Currently, there are two common feature selection strategies: the single-view based feature selection (SVFS) strategy and the concatenation-based feature selection (CFS) strategy.

SVFS selects one special set from multiple types of feature sets. For a better illustration, an example is displayed in **Figure 1**, where the feature 1 to feature N in **Figure 1B** represent N types of feature set, and the feature 1 is selected as the best feature set according to a given strategy (see **Figure 1C**). Due to its simplicity, this strategy has been widely applied in current studies. However, one type of feature set only reflects the characteristics from a single view, ignoring the personalized differences between FC sequences, thus could not comprehensively capture the subtle damage of brain functional tissues caused by neuropsychiatric diseases.

We further give a simple example in **Figure 2**, where $R^P_{3,6}$ denotes the FC sequence between the 3-rd ROI and the 6-th ROI of a given patient, $R^N_{3,6}$ denotes the FC sequence between the 3-rd ROI and the 6-th ROI of a normal control, and the values in the brackets denote the mean and variance of each FC sequence. As can be seen from **Figure 2C1**, in $R_{3,6}$ (i.e., FC sequences between 3-rd ROI and 6-th ROI), the mean displays significant differences between the patient and the normal control, while the variance has no significant differences. On the other hand, in $R_{5,9}$, as shown in **Figure 2C2**, the opposite is true, i.e., the variance largely varies between the patient and the normal control, while the mean has no significant differences. On one hand, solely using the mean as the inputting feature for subsequent classification will not fully tap into the discriminative ability of $R_{5,9}$ since the mean of $R_{5,9}$ is the same in both the patient and the normal control. On the other hand, we notice that only using the variance does not fully characterize the discriminative ability of $R_{3,6}$. Therefore, SVFS methods might fail in fully utilized the discriminative power implied in part of FC sequences.

CFS, as an alternative feature selection and aggregation strategy, concatenates many types of feature set and inputs the resulting set to a target classifier for training. **Figure 1** displays an illustration of this technique. Specifically, all features extracted by various feature extraction methods (**Figure 1B**) are concatenated into a long feature vector (**Figure 1D1**). Next, a feature evaluation method is used to select a few discriminative features (**Figure 1D2**), thereby forming a feature vector with a smaller size for the subsequent classification. However, the main drawback of CFS is that it may result in a sequence selection where for a few FC sequences, many types of features are selected for the subsequent classification, in contrast, for other FC sequences, not any features are selected. As can be seen in **Figure 1D2**, two types of features (i.e., $f_{2,1}$ and $f_{2,2}$) are selected from FC sequence 2, while no features of FC sequence 3 are included. Such strategy may lead to the issue that the selected smaller feature vector contains some redundant information due to the fact that many selected features as they are derived from the same FC sequence. Besides, such unbalanced selection by overlooking the diversity of features across all FC sequences may result in the inability to fully explore the discriminative information of each FC sequence.

It is worth noting that, brain diseases can cause abnormal connections among many pairs of ROIs due to the damage of brain. We believe that the manifestations of connectivity differences among different pair of ROIs are distinctive due to their distinct working mechanisms (Smith et al., 2011), which causes the discriminative fingerprint features from different FC sequences to exist huge difference. As explained above (**Figure 2**), the mean of $R_{3,6}$ is more discriminative but its variance is worse, while the variance of $R_{5,9}$ is more discriminative but its mean is worse, which indicates the mean can reflect the personalized connection characteristics of the 3-rd and 6-th ROIs, and the variance can reflect the personalized connection characteristics

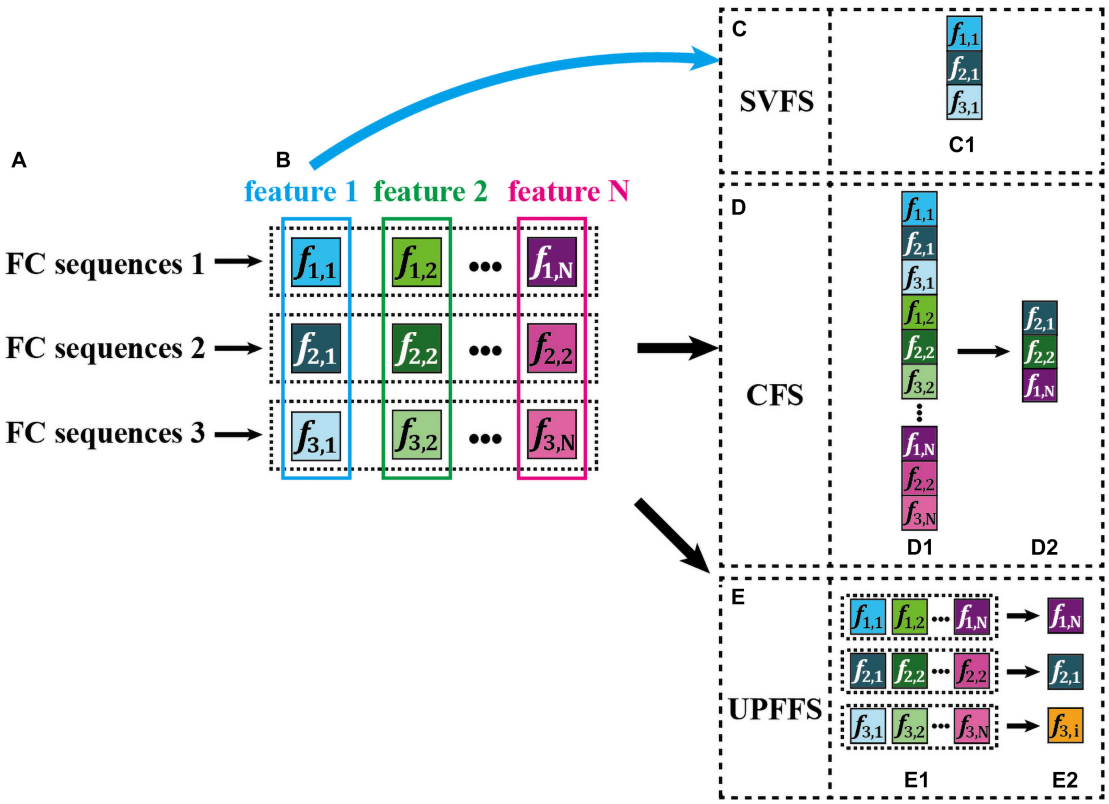


FIGURE 1 | An intuitive explanation for three feature selection strategies. Here, the color saturation is used to encode the feature discriminative power. The higher the saturation, the more discriminative the features. Furthermore, the features with the highest color saturation are displayed in white text for easy viewing. **(A)** Three example FC sequences from D-FCN. **(B)** Multiple features extracted from each FC sequence, i.e., feature 1, feature 2. . . feature N. **(C)** Single-view based feature selection (SVFS) strategy. **(D)** Concatenation based feature selection (CFS) strategy. **(E)** The proposed unit-based personalized fingerprint feature selection (UPFFS) strategy.

of the 5-th and 9-th ROIs. However, as previously described, both traditional CFS and SVFS feature selection strategies cannot fully explore the personalized feature of each FC sequence for more accurate classification of brain diseases.

To address the above issue, we propose a novel unit-based personalized fingerprint feature selection (UPFFS) strategy. Specifically, the FC sequence between any pair of ROIs is regarded as a unit which is handled independently. By using feature extraction methods, we extract multi-view features jointly from each unit, such as feature 1, feature 2, . . ., and feature N as shown in **Figure 1B**. Next, for each unit, the most discriminative one, which is regarded as its fingerprint feature, is selected from multi-view features by using a specific feature evaluation method, as shown in **Figures 1E1,E2**. Last, all the selected features are concatenated as a final feature set for the downstream classification task.

In summary, the main advantages of the proposed strategy are twofold: (1) For different FC sequences, the most discriminative feature types were are distinctive. The proposed UPFFS strategy can better extract personalized features according to the characteristic of each FC sequence, thus solving the single-view problem of SVFS; (2) due to the global damage of disease to human brain, the UPFFS can fully explore the damage information contained in each ROI pair, thereby avoiding the redundant information of CFS caused by the selection bias of the FC sequence, and while fully exploring the discriminative information of each FC sequence.

The rest of this article is organized as follows. In the "Materials and Methods" section, we introduce two baseline strategies (i.e., SVFS and CFS) and provide the details of our proposed unit-based personalized fingerprint feature selection (UPFFS) strategy. In the "Experiment and Results" section, we detail our experimental setting, including network construction, feature selection, classifier construction, and the evaluation measures. In the "Discussion" section, we discuss the impact of different feature evaluation methods, compare our strategy with other benchmarks, and analyze the identified discriminative features. Finally, the "Conclusion" section summarizes the paper.

MATERIALS AND METHODS

Figure 3 illustrates the flow of these three strategies (i.e., SVFS, CFS, and the proposed UPFFS) which will be described in detail in the following subsections. We also display the feature processing pipeline prior to feature selection while the detail of

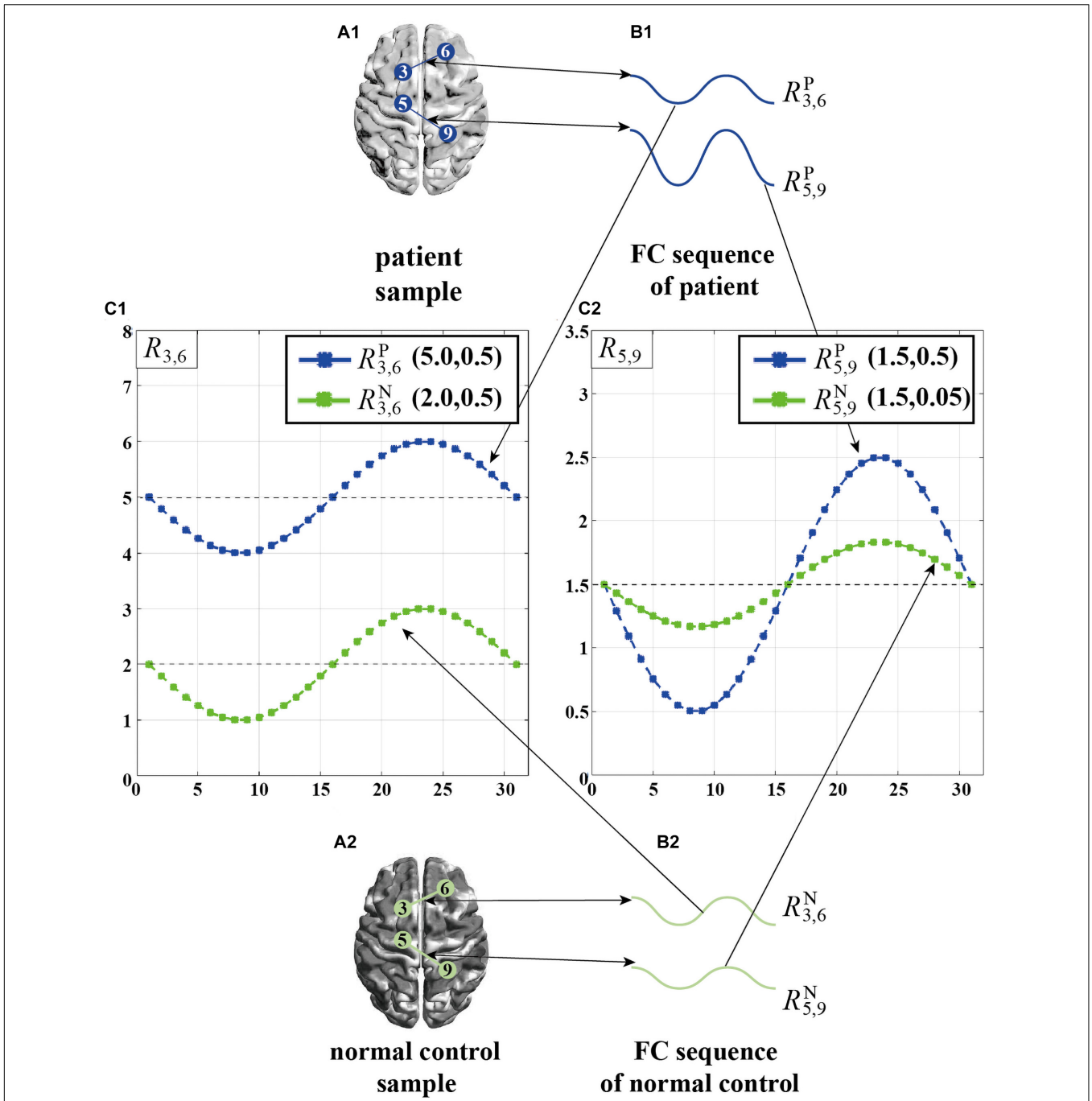


FIGURE 2 | A simple example of the individual difference in the manifestation of connectivity differences between brain ROI pairs. **(A1)** Patient sample. **(A2)** Normal control sample. **(B1)** Two FC sequences from patient sample, $R^{P}_{3,6}$ and $R^{P}_{5,9}$. **(B2)** Two FC sequences from normal control sample, $R^{N}_{3,6}$ and $R^{N}_{5,9}$. **(C1)** Display of $R^{P}_{3,6}$ and $R^{N}_{3,6}$. **(C2)** Display of $R^{P}_{5,9}$ and $R^{N}_{5,9}$.

the feature processing will be introduced in the "Experiment and Results" section.

For ease of description, we define the variables that will be involved in the following subsections: (1) $\rho_{i,j}(k)$ denotes the correlation between the i-th ROI and the j-th ROI in k-th time window; (2) $\boldsymbol{\rho}_{i,j} = \{\rho_{i,j}(1), \rho_{i,j}(2), \ldots, \rho_{i,j}(K)\}$ $(i, j = 1, 2, \ldots, M)$ denotes the FC sequence between the i-th ROI and the j-th ROI where M is the number of ROIs and K is the number of time windows; (3) $f_{i,j}(d)$ denotes the feature extracted from $\boldsymbol{\rho}_{i,j}$ with the d-th feature extraction method; (4) $\boldsymbol{F}_d = [f_{i,j}(d)]_{1<i,j<M}$ $(d = 1, 2, 3, \ldots, D)$ aggregates all d-th features in $\boldsymbol{\rho}_{i,j}$ across all pairs of ROIs i and j.

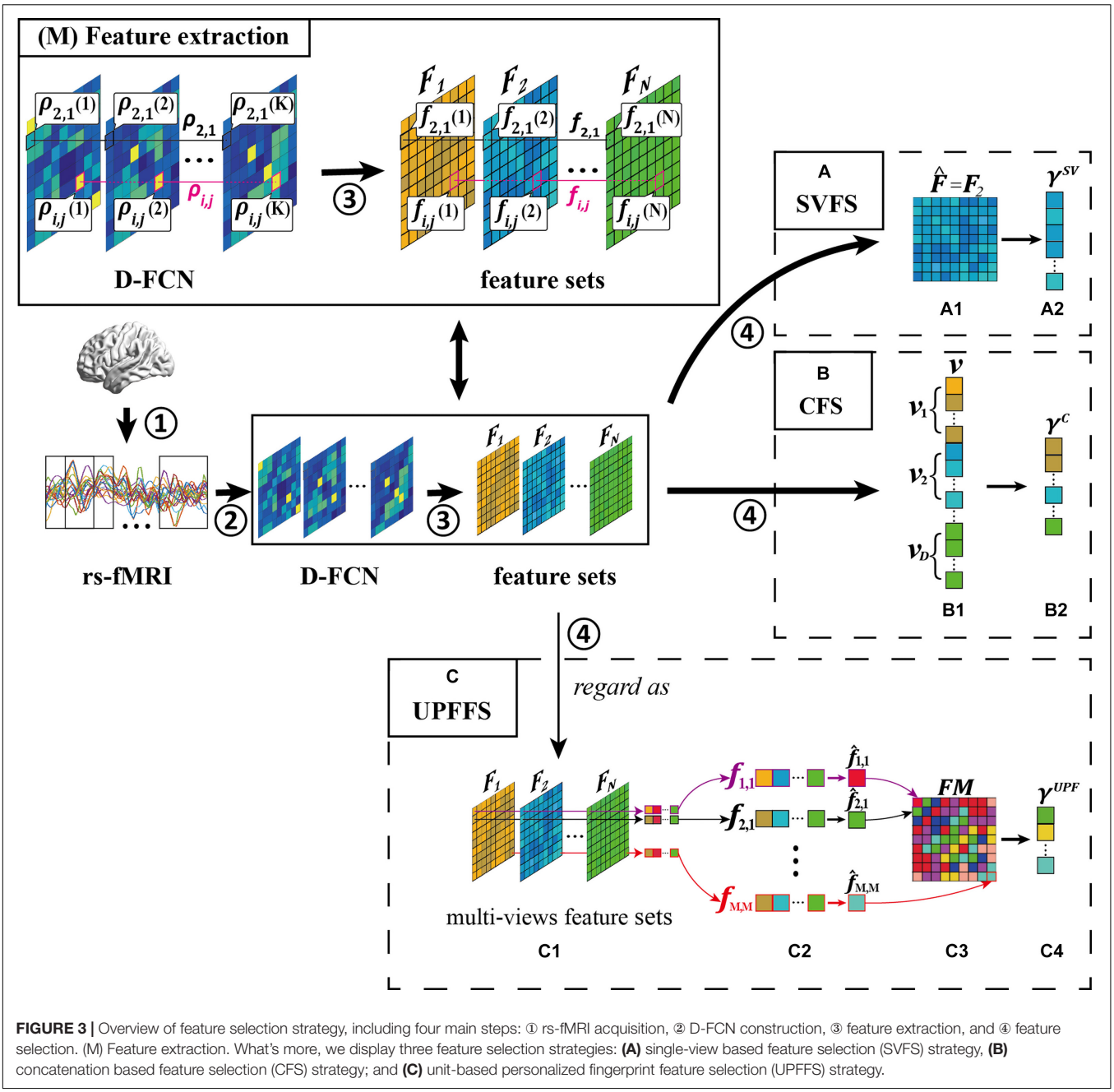


FIGURE 3 | Overview of feature selection strategy, including four main steps: ① rs-fMRI acquisition, ② D-FCN construction, ③ feature extraction, and ④ feature selection. (M) Feature extraction. What's more, we display three feature selection strategies: **(A)** single-view based feature selection (SVFS) strategy, **(B)** concatenation based feature selection (CFS) strategy; and **(C)** unit-based personalized fingerprint feature selection (UPFFS) strategy.

Traditional Feature Selection Strategies

Single-View Based Feature Selection Strategy

The main idea of single-view based feature selection (SVFS) is to select the best-performing type of feature set for the target classification task. Specifically, SVFS firstly tests the performance of each feature set $\boldsymbol{F}_d$ with a specific criterion ($d = 1, 2, \ldots, D$, where D is the number of feature extraction methods as motioned above), such as the classification accuracy, and then the best-performing feature set $\hat{\boldsymbol{F}}$ is selected according to the following definition:

$$\hat{F} = \text{opt}\{F_d\}_{d=1}^{D} \quad (1)$$

where $\text{opt}\{\boldsymbol{F}_d\}_{d=1}^{D}$ denotes the selection of the best one from all types of feature sets.

Next, in order to remove the redundant features and preserve the discriminative features that are most likely relevant to disease, SVFS utilizes the feature evaluation method (e.g., the two-sample t-test between normal control and patient subjects) in the next step and the features with better evaluation results are retained to form the final feature set $\boldsymbol{\gamma}^{SV}$.

Figure 3A shows an example of SVFS. In this example, feature set 2 is selected as the optimal feature set in **Figure 3A1**, (i.e., $\hat{\boldsymbol{F}} = \boldsymbol{F}_2$). Then, in **Figure 3A2**, the discriminative features are selected to form the final feature set $\boldsymbol{\gamma}^{SV}$.

In this way, the feature set that are most between-class discriminative can be selected. However, the personality differences of each FC sequence are ignored which may hinder the identification of the disease-altered brain connectivity patterns in a comprehensive manner, because some FC sequences may show discrimination information in other feature sets (e.g., $\boldsymbol{F}_1, \boldsymbol{F}_3, \ldots, \boldsymbol{F}_D$) instead of $\boldsymbol{F}_2$.

Concatenation Based Feature Selection Strategy

The main idea of Concatenation Based Feature Selection (CFS) is to firstly concatenate many types of features as a big feature set and then filter out some discriminative features as the final feature set for classification. Specifically, in the first step, the features from each feature set $\boldsymbol{F}_d$ $(d = 1, 2, \ldots, D)$ are vectorized as a vector $\boldsymbol{v_d}$. Next, all $\boldsymbol{v_d}$ are concatenated into a long vector $\boldsymbol{v}$ and the discriminative features are selected from $\boldsymbol{v}$ with the feature evaluation method to form the final feature set $\boldsymbol{\gamma}^{\boldsymbol{C}}$.

Figure 3B is an example of CFS where the vectors $\boldsymbol{v_d}$ from each feature set $\boldsymbol{F_d}$ are concatenated into a long vector in **Figure 3B1**, and in **Figure 3B2**, the discriminative features are selected to form the final feature set $\boldsymbol{\gamma}^{\boldsymbol{C}}$.

Therefore, the discriminative information from all feature sets can be explored in such a way. However, the information redundancy that may be caused by the selected features in the same FC sequence and the omission of some FC sequences will affect the classification results.

Unit-Based Personalized Fingerprint Feature Selection Strategy

In this subsection, we introduce a novel unit-based personalized fingerprint feature selection (UPFFS) strategy, which takes an adaptive approach in extracting the personalized fingerprint feature for each FC sequence.

Specifically, a FC sequence $\boldsymbol{\rho}_{i,j} = \left\{\rho_{i,j}(1), \rho_{i,j}(2), \ldots, \rho_{i,j}(K)\right\}\left(i, j = 1, 2, \ldots, M\right)$ is regarded as a unit and the multiple types of features from a unit are combined as a feature vector $\boldsymbol{f}_{i,j}$ (i.e., $\boldsymbol{f}_{i,j} = \left\{f_{i,j}(1), f_{i,j}(2), \ldots, f_{i,j}(K)\right\}$ where $f_{i,j}(1)$ denotes the first type of feature from $\boldsymbol{\rho}_{i,j}$), as shown in **Figure 3C1**. Next, the feature evaluation method is performed for each $\boldsymbol{f}_{i,j}$ internally. In this way, the most discriminative personalized fingerprint feature $\hat{f}_{i,j}$ (**Figure 3C2**) with the best evaluation result will be selected from each unit using the following formula:

$$\hat{f}_{i,j} = \mathrm{opt}\{f_{i,j}(d)\}_{d=1}^{D} \tag{2}$$

where the $\mathrm{opt}\{f_{i,j}(d)\}_{d=1}^{D}$ denotes the selection of the best evaluation result from $f_{i,j}$.

For the whole D-FCN, the fingerprint feature matrix $\boldsymbol{FM}$ (**Figure 3C3**) can be composed with the fingerprint features $\hat{f}_{i,j}$ of all FC sequences according to the following definition:

$$\boldsymbol{FM} = \left[\hat{f}_{i,j}\right]_{1<i,j<M} \tag{3}$$

The features with better evaluation results than the default threshold will also be preserved to form a final fingerprint feature set $\boldsymbol{\gamma}^{\boldsymbol{UPF}}$ (**Figure 3C4**) to avoid the impact of units which are less relevant to the disease.

This approach can not only solve the single-view problem of SVFS by considering multiple feature sets simultaneously, but also fully take into account the personalized features of each FC sequence which avoids the feature selection imbalance in CFS.

EXPERIMENT AND RESULTS

To evaluate the effectiveness of the framework, we perform classification on the ASD subjects and normal controls. Specifically, we construct the traditional SW-D-FCN which is also called low-order dynamic functional connectivity networks (Lo-D-FCN), and the SW-D-FCN based high-order dynamic functional connectivity networks (Ho-D-FCN) on the rs-fMRI data set from the Autism Brain Imaging Data Exchange (ABIDE) database (Di Martino et al., 2014), apply seven feature extraction methods based on the central moment method, and then input the extracted features into a linear support vector machine (SVM) for linear fusion. In particular, the ASD data set and data preprocessing are same as those in a recent study (Wee et al., 2016).

Comparison of Feature Selection Strategies Based on Lo-D-FCN

Lo-D-FCN Construction. In the construction of Lo-D-FCN, the entire rs-fMRI time series are firstly divided into K overlapping sub-windows by a sliding window with a prefixed window length W and a step size S (**Figure 4A1**). Next, for each sub-window, the correlations between the rs-fMRI time series from each pair of ROIs are calculated to create the sub Lo-FCN. As the example shown in **Figures 4A2–B2**, in the sub-window 2, $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ denote the rs-fMRI time series from the i-th and the j-th ROI respectively, and $\rho_{i,j}(2)$, as a component of sub Lo-FCN 2 (i.e., the second subnetwork of Lo-D-FCNs), denote the Pearson's correlation (PC) of $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$. Repeating the above process, a series of sub Lo-FCN can be constructed which is called Lo-D-FCN (**Figure 4B1**). Obviously, the Lo-D-FCN can quantify the correlation change between each pair of ROIs in total scanning time.

Feature Extraction. In this experiment, we utilized seven center-distance feature extraction methods for extracting the dynamic variation of FC among multiple ROIs along the scanning time (Zhao et al., 2020). In particular, for Lo-D-FCN, the FC sequence of the i-th and the j-th ROI are defined as $\boldsymbol{\rho}_{i,j} = \left[\rho_{i,j}(1), \rho_{i,j}(2), \ldots, \rho_{i,j}(K)\right]$. $\boldsymbol{\rho}_{i,j}$ reflects the FC dynamic changes of the i-th and the j-th ROI along the scanning timeline, which can be quantified by calculating its center-distance features. The features from all $\boldsymbol{\rho}_{i,j}$ can form a series of feature matrices.

Classifier Learning. The important features are selected using the traditional SVFS, CFS, and the proposed UPFFS strategies respectively from the set of all feature matrices. It is noteworthy that we utilize the t-test as the feature evaluation method in this experiment. Besides, we employ a sixfold cross-validation (CV) strategy to perform experiments and train the linear support

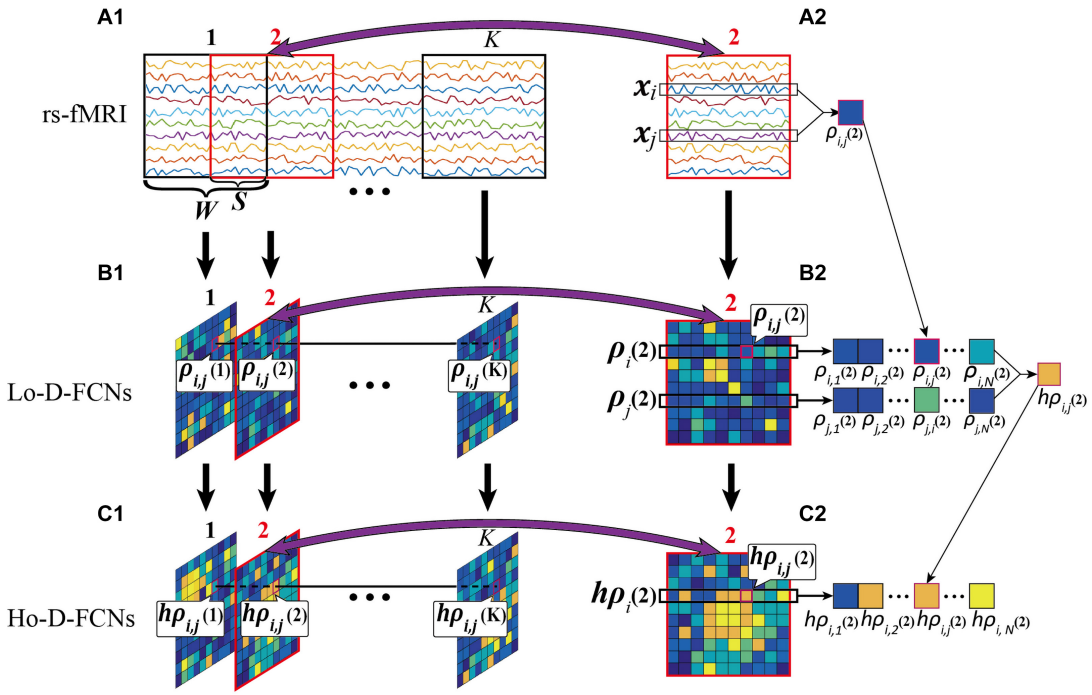


FIGURE 4 | Flow chart of constructing Lo-D-FCN and Ho-D-FCN, where **(A1)** is the rs-fMRI series, **(A2)** is the second rs-fMRI subseries based on a sliding window, **(B1)** denotes the Lo-D-FCNs, **(B2)** denotes the second subnetwork of Lo-D-FCNs, **(C1)** denotes the Ho-D-FCNs, and **(C2)** denotes the second subnetwork from Ho-D-FCNs.

vector machines (SVMs) (Cortes and Vapnik, 1995) for ASD classification tasks. For instance, one subject is selected as the testing dataset for each CV, while the other 5 subsets are used as the training dataset, and further, the nested fivefold CV is utilized on these 5 subsets to adjust the p-value in t-test model, as well as gamma (γ) and alpha (α) in SVM (classification learning and fusion parameters). In particular, the type of most discriminative feature of each FC sequence in UPFFS is determined with the training dataset.

For fair comparison, the sixfold CV was repeated 10 times and the samples are randomly partitioned each time differently to reduce the bias due to random seed selection and the optimal parameters are limited in the following range: p-value$\in [0.01 : 0.01 : 0.1]$, $\gamma \in \left[2^{-5}, 2^{-4}, \ldots, 2^{5}\right]$ and $\alpha \in [0.1 : 0.1 : 0.9]$.

Classification Results. Similar to previous studies (Zhao et al., 2017), we use classification accuracy (ACC), sensitivity or true positive rate (TPR), specificity or true negative rate (TNR), positive predictive value (PPV), negative predictive value (NPV), F1 score[1] as performance indicators to quantify the performance of different feature selection strategies. Notably, in the classification work of this experiment, we consider ASD patients as positive and normal controls as negative.

Table 1 reports the best results of ASD identification with SVFS, CFS, and our UPFFS strategy in Lo-D-FCN where $SVFS_{Lo}$ denotes the classification model using Lo-D-FCN and SVFS strategy. Specifically, for Lo-D-FCN construction, we select $W = 60$, $S = 2$ (e.g., the sliding-window length W equals to 60 and step size S equals to 2) as the optimal construction parameters according to the previous study (Zhao et al., 2020). The best results are highlighted in bold.

From **Table 1**, we note that: (1) The proposed UPFFS strategy achieves better results than when using one of the seven center-distance methods alone (SVFS) or simple concatenation features (CFS); (2) the CFS just marginally boosts the classification accuracy compared with SVFS. The possible reason for such results is the unbalanced selection of the features from different FC sequences as previously mentioned in the "Introduction" section, and the UPFFS achieves better results by combining multiple types of features while avoiding this problem.

Comparison of Feature Selection Strategies Based on Ho-D-FCN

Ho-D-FCN Construction. To further study the impact of different SW-D-FCN artifacts on the classification performance, we applied the above three strategies on the "correlation's correlation" (Morris and Rekik, 2017; Soussia and Rekik, 2018) based Ho-D-FCN which is constructed by the sub Ho-FCN calculated from each sub Lo-FCN to extract higher-level interactions between ROIs. Specifically, in each Lo-FCN the correlation between one ROI and all other ROIs is regarded as a sequence, and the correlation between these sequences is calculated to form a Ho-FCN.

[1]https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Figures 4B2–C2 shows an example of constructing the sub Ho-FCN 2 where $\boldsymbol{\rho}_i(2)$ denotes the sequence of the correlations between the i-th ROI and the other ROIs in sub-window 2, and $h\rho_{i,j}(2)$, as a component of this sub Ho-FCN, denotes the correlation of $\boldsymbol{\rho}_i(2)$ and $\boldsymbol{\rho}_j(2)$. Repeating the above process, as shown in **Figure 4C1**, a series of sub Ho-FCN can be constructed which is called Ho-D-FCN. Obviously, the Ho-D-FCN can reflect the interaction pattern of temporal correlations from different ROI pairs over the whole scan time.

Classification Results. For Ho-D-FCN construction, we select $W = 60$, $S = 10$ as the construction parameters according to the previous research (Zhao et al., 2020). Since feature extraction and classifier learning are consistent with those above, we will not repeat them here. **Table 2** displays the classification results where $SVFS_{Ho}$ denotes the classification model with Ho-D-FCN and SVFS strategy. The best results are also highlighted in bold.

It can be seen that the results of Ho-D-FCN are similar to those of Lo-D-FCN despite the modest improvement in accuracy. Specifically, the proposed UPFFS performs significantly better than these two traditional strategies in Ho-D-FCN, and CFS also slightly improves the classification performance compared with SVFS, which further supports our preliminary hypothesis.

Classification Performance Based on the Fusion of Lo-D-FCN and Ho-D-FCN

There are complex connections inside the brain, thus it is difficult to fully capture the relationship between various regions through a single type of FCN. In order to further improve the classification performance, it is a general practice to combine the Lo-D-FCNs and Ho-D-FCNs by linear fusion of SVM ensemble decision scores (Zhao et al., 2020). In this paper, we fuse the SVM scores from two D-FCNs using different feature selection strategies.

TABLE 1 | ASD classification results with different feature selection strategies using Lo-D-FCN.

Model	ACC (%)	TPR (%)	TNR (%)	PPV (%)	NPV (%)	F1 (%)
$SVFS_{Lo}$	70.4 ± 0.14	70.0 ± 0.45	70.8 ± 0.15	69.7 ± 0.11	71.4 ± 0.55	69.7 ± 0.20
CFS_{Lo}	70.9 ± 0.15	69.6 ± 0.48	72.1 ± 0.23	70.5 ± 0.14	71.5 ± 0.21	69.9 ± 0.22
$UPFFS_{Lo}$	**73.6 ± 0.06**	**70.7 ± 0.12**	**76.4 ± 0.11**	**74.2 ± 0.09**	**73.2 ± 0.06**	**72.3 ± 0.07**

TABLE 2 | ASD classification results with different feature selection strategies using Ho-D-FCN.

Model	ACC (%)	TPR (%)	TNR (%)	PPV (%)	NPV (%)	F1 (%)
$SVFS_{Ho}$	70.7 ± 0.04	71.1 ± 0.12	70.2 ± 0.05	69.6 ± 0.03	71.8 ± 0.06	70.3 ± 0.05
CFS_{Ho}	71.5 ± 0.17	71.3 ± 0.38	71.7 ± 0.14	70.7 ± 0.14	72.5 ± 0.24	70.9 ± 0.22
$UPFFS_{Ho}$	**74.0± 0.10**	**73.8 ± 0.25**	**74.3± 0.12**	**73.3± 0.09**	**74.9± 0.15**	**73.5± 0.13**

TABLE 3 | Fusion result of different feature selection strategies.

Model	ACC (%)	Promote (Lo/Ho)	TPR (%)	TNR (%)	PPV (%)	NPV (%)	F1 (%)
$SVFS_{Lo} + SVFS_{Ho}$	71.4 ± 0.23	1.0/0.7	73.1 ± 0.16	69.8 ± 0.14	69.9 ± 0.10	73.1 ± 0.12	71.4 ± 0.11
$CFS_{Lo} + CFS_{Ho}$	72.8 ± 0.07	1.9/1.3	72.0 ± 0.36	73.6 ± 0.14	72.3 ± 0.06	73.5 ± 0.14	72.1 ± 0.13
$UPFFS_{Lo} + UPFFS_{Ho}$	**76.0 ± 0.10**	**2.4/2.0**	**74.4 ± 0.31**	**77.5 ± 0.20**	**76.0 ± 0.13**	**76.1 ± 0.14**	**75.1 ± 0.14**

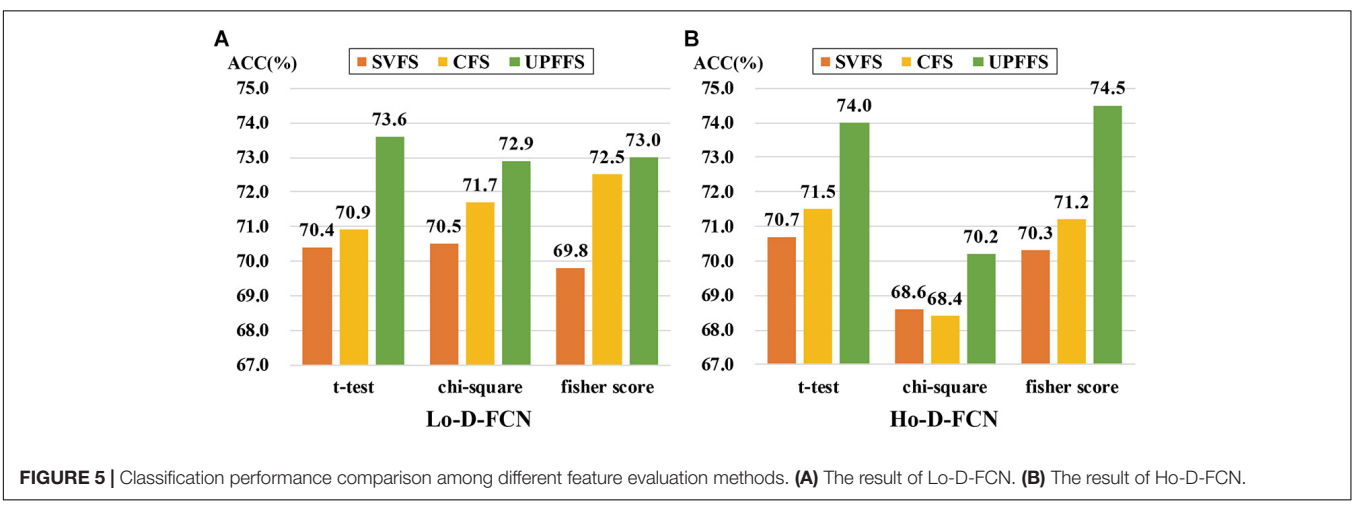


FIGURE 5 | Classification performance comparison among different feature evaluation methods. **(A)** The result of Lo-D-FCN. **(B)** The result of Ho-D-FCN.

Specifically, each SVM classifier can output a decision score indicating the probability of a subject belonging to a class. So we calculate the weighted average of the decision scores from two SVM classifiers to obtain the classification results, where the weight tuned for each SVM is determined using a nested CV. The final classification results are shown in **Table 3**, where the symbol " + " represents the fusion, for example, $SVFS_{Lo} + SVFS_{Ho}$ represents the fusion of Lo-D-FCN and Ho-D-FCN under the SVFS. We also highlight the best results with a bold font in **Table 3**.

In addition to the indicators above, we also display the improvement of fusion results compared to the single D-FCN in classification accuracy with the Promote(Lo/Ho) indicator (e.g., 2.4/2.0 represents that the fusion network improved 2.4 compared to Lo-D-FCN and 2.0 compared to Ho-D-FCN in the classification accuracy). Based on **Table 3**, we derive the following conclusions: (1) the feature fusion result consistently produces better results than features derived from a single network, i.e., the promote indicators are all greater than 0, which is similar to the conclusions of previous studies (Zhao et al., 2020); (2) our UPFFS strategy has better results and greater improvement compared with other strategies, which indicates that our feature selection framework can select more complementary discriminative features from Lo-D-FCN and Ho-D-FCN, thus the SVM classifier has a better learning and generalizability to unseen samples.

TABLE 4 | Abbreviations of ROIs selected from Lo-D-FCN and Ho-D-FCN.

Abbreviation	ROI name	Abbreviation	ROI name
PreCG	Precentral gyrus	ROBsup	Orbitofrontal cortex (middle)
IFGoperc	Inferior frontal gyrus (opercular)	IFGtrian	Inferior frontal gyrus (triangular)
SMA	Suplementary motor area	SFGmed	Superior frontal gyrus (medial)
ORBsupmed	Orbitofrontal cortex (medial)	REC	Rectus gyrus
PCG	Posterior cingulate gyrus	HIP	Hippocampus
PHG	ParaHippocampal gyrus	CAL	Calcarine cortex
LING	Lingual gyru	IPL	Inferior parietal lobul
SMG	Supramarginal gyrus	ANG	Angular gyrus
PAL	Pallidum	TPOmid	Temporal pole (middle)
IV-V-VER	Lobule IV, V of vermis	VI-VER	Lobule VI of vermis

DISCUSSION

Effect of Different Feature Evaluation Method

The feature evaluation method is an important part of our UPFFS strategy. In order to investigate the effect of the feature evaluation method on the classification performance, we further utilize chi-square and fisher score as our feature evaluation method respectively to replace the *t*-test, then perform classification experiments following the same steps. In particular, the threshold for alternative feature evaluation method is also adjusted by nested CV, where the threshold range of chi-square is set to $[0.1 : 0.1 : 1]$ and the threshold range of fisher score is $[0.01 : 0.005 : 0.05]$ according to previous studies. The experimental results are shown in **Figure 5**.

It can be seen from **Figure 5** that UPFFS always has a better result than traditional strategies with different feature evaluation methods, which further shows the outperformance of the UPFFS strategy. On the other hand, we empirically found that the performances are sensitive to the feature evaluation methods. For instance, the ASD classification results of chi-square in Ho-D-FCN (**Figure 5B**) are lower than the others. Thus, in

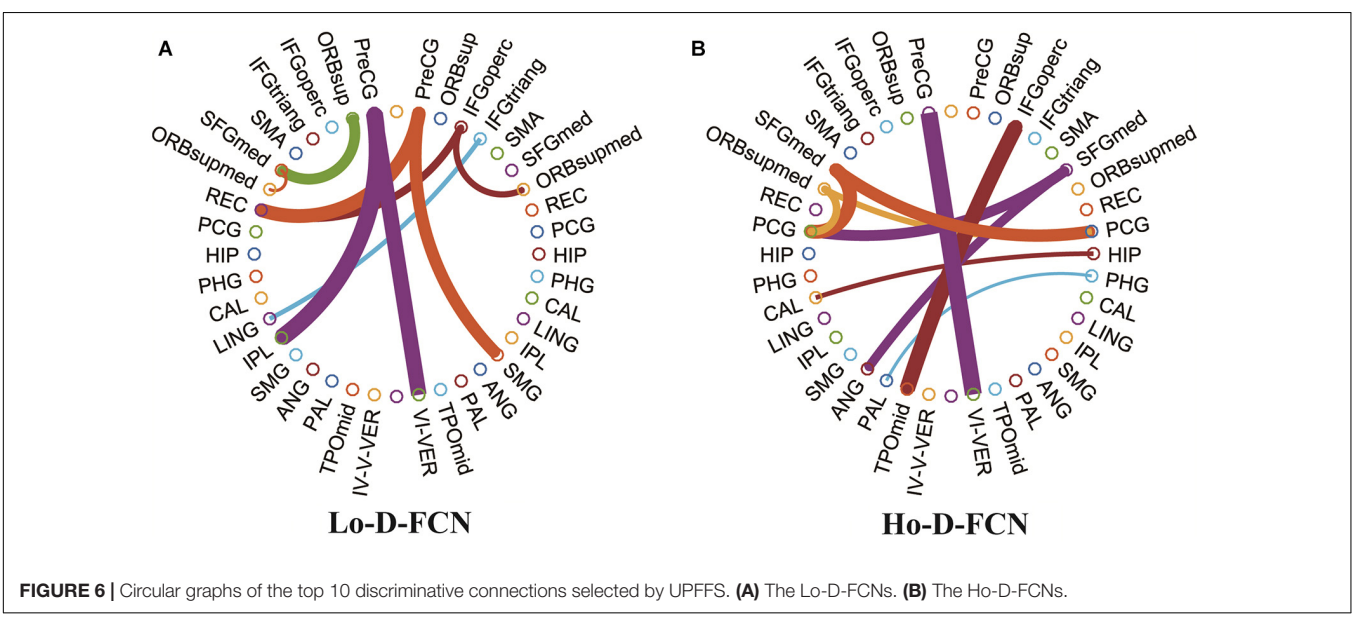


FIGURE 6 | Circular graphs of the top 10 discriminative connections selected by UPFFS. **(A)** The Lo-D-FCNs. **(B)** The Ho-D-FCNs.

practice, the feature evaluation methods of UPFFS should be utilized carefully depending on the dataset in hand and the target classification task.

Top Discriminative Feature

For exploring the internal reasons that the UPFFS strategy improves the ASD classification performance from a physiological aspect, we identify the set of the most discriminative features. Specifically, we count the frequency at which features are selected within the sixfold cross-validation and quantify the discriminative ability of one feature with its frequency. The higher its frequency, the more discriminative it is regarded.

Figure 6 visualizes the top 10 most discriminative features selected by UPFFS strategy in Lo-D-FCN (**Figure 6A**) and Ho-D-FCN (**Figure 6B**) respectively with circular graphs, where a node denotes an ROI and a line between the two nodes denotes a connection feature which represents the correlation between two ROIs (Krzywinski et al., 2009). **Table 4** lists the abbreviations of brain regions in **Figure 6**.

From the results shown in **Figure 6** and **Table 4**, we find that: (1) the corresponding ROIs of the top discriminative connection features selected by the UPFFS strategy include the superior frontal gyrus, orbitofrontal cortex, hippocampus, calcarine cortex, angular gyrus, etc., which are associated with visual processing, social cognition, and emotional expression. These findings are consistent with previous studies (Di Martino et al., 2009; Scherf et al., 2015). In particular, superior frontal gyrus (medial) (Goldberg et al., 2006; Sato et al., 2012), orbitofrontal cortex (medial) (Ye et al., 2014), hippocampus (Ha et al., 2015), calcarine cortex (Libero et al., 2014; Perkins et al., 2015), angular gyrus (Goldberg et al., 2006) have been reported as potential biomarkers for ASD identification. (2) Most of the selected discriminative features connect the brain lobe of transhemisphere rather than the same hemisphere, which indicates that the UPFFS strategy can identify abnormal distribution patterns over the whole brain. (3) The discriminative features of Lo-D-FCNs, and Ho-D-FCNs are obviously different, indicating that these two networks can contain complementary information for ASD and that our UPFFS strategy is able to capture such rich and personalized information.

CONCLUSION

In this paper, we proposed a new feature selection strategy, called UPFFS, which regarded the FC sequence between any pair of brain ROIs as a unit and extracts the personalized fingerprint feature from each unit, thus reflecting the discriminative information in SW-D-FCN more comprehensively. We evaluated the performance of the proposed UPFFS and two traditional approaches on two types of SW-D-FCN. The experimental results have shown that: (1) UPFFS always has better ability to extract discriminative information compared with the two traditional approaches; (2) UPFFS is able to extract features that are unique to the network, making features from different networks more complementary and causing a higher improvement in classification result for the fusion of two networks; (3) we also found that the top selected discriminative brain regions by UPFFS are related to visual processing, social cognition, and emotional expression which is in line with previous research results and further validates the effectiveness of UPFFS.

Although we proposed this strategy for SW-D-FCN, this idea can be utilized in other fields requiring personalized feature identification, and it should be indicated that the feature evaluation method have a certain influence on the classification results. In our future work, we will apply the UPFFS to other types of "omic" datasets such as genomics for underpinning biological biomarker in a variety of disorders.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported in part by the National Natural Science Foundation of China (61773244, 82001775, 61772319, 61873177, 61972235, and 61976125), Yantai Key Research and Development Program of China (2017ZH065 and 2019XDHZ081), Shandong Provincial Key Research and Development Program of China (2019GGX101069), and Doctoral Scientific Research Foundation of Shandong Technology and Business (BS202016).

REFERENCES

Biao, J., Wee, C. Y., Shen, D., and Zhang, D. (2016). Hyper-connectivity of functional networks for brain disease diagnosis. *Med. Image Anal.* 32, 84–100. doi: 10.1016/j.media.2016.03.003

Chen, X., Zhang, H., Gao, Y., Wee, C. Y., Li, G., Shen, D., et al. (2016). High-order resting-state functional connectivity network for MCI classification. *Hum. Brain Mapp.* 37, 3282–3296. doi: 10.1002/hbm.23240

Chen, X., Zhang, H., Zhang, L., Shen, C., Lee, S. W., and Shen, D. (2017). Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification. *Hum. Brain Mapp.* 38, 5019–5034. doi: 10.1002/hbm.23711

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine Learn.* 20 273–297.

Cullen, K. R., Westlund, M. K., Klimes-Dougan, B., Mueller, B. A., Houri, A., Eberly, L. E., et al. (2014). Abnormal amygdala resting-state functional

connectivity in adolescent depression. *JAMA Psychiatry* 71, 1138–1147. doi: 10.1001/jamapsychiatry.2014.1087

Damaraju, E., Allen, E. A., Belger, A., Ford, J. M., McEwen, S., Mathalon, D. H., et al. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage* 5, 298–308. doi: 10.1016/j nicl.2014.07.003

Di Martino, A., Shehzad, Z., Kelly, C., Roy, A. K., Gee, D. G., Uddin, L. Q., et al. (2009). Relationship between cingulo-insular functional connectivity and autistic traits in neurotypical adults. *Am. J. Psychiatry* 166, 891–899. doi: 10. 1176/appi.ajp.2009.08121894

Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78

Du, Y., Fu, Z., and Calhoun, V. (2018). Classification and prediction of brain disorders using functional connectivity: promising but challenging. *Front. Neurosci.* 12:525. doi: 10.3389/fnins.2018. 00525

Fornito, A., Zalesky, A., and Breakspear, M. (2015). The connectomics of brain disorders. *Nat. Rev. Neurosci.* 16, 159–172. doi: 10.1038/nrn3901

Goldberg, I. I., Harel, M., and Malach, R. (2006). When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 50, 329–339. doi: 10.1016/j.neuron.2006.03.015

Greicius, M. D., Flores, B. H., Menon, V., Glover, G. H., Solvason, H. B., Kenna, H., et al. (2007). Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiatry* 62, 429–437. doi: 10.1016/j.biopsych.2006. 09.020

Ha, S., Sohn, I. J., Kim, N., Sim, H. J., and Cheon, K. A. (2015). Characteristics of brains in autism spectrum disorder: structure, function and connectivity across the lifespan. *Exp. Neurobiol.* 24, 273–284.doi: 10.5607/en.2015.24. 4.273

Huang, H., Liu, X., Jin, Y., Lee, S. W., Wee, C. Y., and Shen, D. (2018). Enhancing the representation of functional connectivity networks by fusing multi-view information for autism spectrum disorder diagnosis. *Hum. Brain Mapp.* 40, 833–854. doi: 10.1002/hbm.24415

Jia, H., Wu, G., Wang, Q., and Shen, D. (2010). ABSORB: atlas building by self-organized registration and bundling. *NeuroImage* 51, 1057–1070. doi: 10.1016/ j.neuroimage.2010.03.010

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Libero, L. E., Maximo, J. O., Deshpande, H. D., Klinger, L. G., Klinger, M. R., and Kana, R. K. (2014). The role of mirroring and mentalizing networks in mediating action intentions in autism. *Mol. Autism* 5:50. doi: 10.1186/2040-2392-5-50

Liu, F., Wang, Y., Li, M., Wang, W., Li, R., Zhang, Z., et al. (2016). Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic-clonic seizure. *Hum. Brain Mapp.* 38, 957–973. doi: 10.1002/ hbm.23430

Liu, Y., Wang, K., Yu, C., He, Y., Zhou, Y., Liang, M., et al. (2008). Regional homogeneity, functional connectivity and imaging markers of Alzheimer's disease: a review of resting-state fMRI studies. *Neuropsychologia* 46, 1648–1656. doi: 10.1016/j.neuropsychologia.2008.01.027

Luan, H., Qi, F., Xue, Z., Chen, L., and Shen, D. (2008). Multimodality image registration by maximization of quantitative–qualitative measure of mutual information. *Pattern Recognition* 41, 285–298. doi: 10.1016/j.patcog.2007.04. 002

Morris, C., and Rekik, I. (2017). "Autism spectrum disorder diagnosis using sparse graph embedding of morphological brain networks," in *Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics: First International Workshop, GRAIL 2017, 6th International Workshop, MFCA 2017, and Third International Workshop, MICGen 2017, Held in Conjunction with MICCAI 2017* eds M. J. Cardoso, and T. Arbel. (Berlin: Springer)12–20. doi: 10.1007/978-3-319-67675-3_2

Perkins, T. J., Bittar, R. G., McGillivray, J. A., Cox, I. I., and Stokes, M. A. (2015). Increased premotor cortex activation in high functioning autism during action observation. *J. Clin. Neurosci.* 22, 664–669. doi: 10.1016/j.jocn.2014. 10.007

Qiao, L., Zhang, L., Chen, S., and Shen, D. (2018). Data-driven graph construction and graph learning: a review. *Neurocomputing* 312, 336–351. doi: 10.1016/j. neucom.2018.05.084

Sato, W., Toichi, M., Uono, S., and Kochiyama, T. (2012). Impaired social brain network for processing dynamic facial expressions in autism spectrum disorders. *BMC Neurosci.* 13:99. doi: 10.1186/1471-2202-13-99

Scherf, K. S., Elbich, D., Minshew, N., and Behrmann, M. (2015). Individual differences in symptom severity and behavior predict neural activation during face processing in adolescents with autism. *NeuroImage Clin.* 7, 53–67. doi: 10.1016/j.nicl.2014.11.003

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for FMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Soussia, M., and Rekik, I. (2018). Unsupervised manifold learning using high-order morphological brain networks derived From T1-w MRI for autism diagnosis. *Front. Neuroinform.* 12:70. doi: 10.3389/fninf.2018. 00070

Van Den Heuvel, M. P., and Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534. doi: 10.1016/j.euroneuro.2010. 03.008

Wee, C. Y., Yang, S., Yap, P. T., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2015). Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. *Brain Imaging Behav.* 10, 342–356. doi: 10.1007/s11682-015-9408-2

Wee, C. Y., Yap, P. T., and Shen, D. (2016). Diagnosis of autism spectrum disorders using temporally distinct resting-state functional connectivity networks. *CNS Neurosci. Ther.* 22, 212–219. doi: 10.1111/cns.12499

Wu, G., Jia, H., Wang, Q., and Shen, D. (2011). SharpMean: groupwise registration guided by sharp mean image and tree-based registration. *NeuroImage* 56, 1968–1981. doi: 10.1016/j.neuroimage.2011.03.050

Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., and Shen, D. (2013). "Unsupervised deep feature learning for deformable registration of MR brain images," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Berlin: Springer), 649–656. doi: 10.1007/978-3-642-40763-5_80

Yaesoubi, M., Allen, E. A., Miller, R. L., and Calhoun, V. D. (2015). Dynamic coherence analysis of resting fMRI data to jointly capture state-based phase, frequency, and time-domain information. *Neuroimage* 120, 133–142. doi: 10. 1016/j.neuroimage.2015.07.002

Ye, A. X., Leung, R. C., Schäfer, C. B., Taylor, M. J., and Doesburg, S. M. (2014). Atypical resting synchrony in autism spectrum disorder. *Hum. Brain Mapp.* 35, 6049–6066. doi: 10.1002/hbm.22604

Zhang, H., Chen, X., Shi, F., Li, G., Kim, M., Giannakopoulos, P., et al. (2016). Topographical information-based high-order functional connectivity and its application in abnormality detection for mild cognitive impairment. *J. Alzheimers Dis.* 54, 1095–1112. doi: 10.3233/jad-16 0092

Zhang, X., Hu, B., Ma, X., and Xu, L. (2015). Resting-state whole-brain functional connectivity networks for MCI classification using L2-regularized logistic regression. *IEEE Trans. Nanobiosci.* 14, 237–247. doi: 10.1109/tnb.2015. 2403274

Zhang, Y., Zhang, H., Chen, X., and Shen, D. (2017). Constructing multi-frequency high-order functional connectivity network for diagnosis of mild cognitive impairment. *Connectomics Neuroimaging* 105119, 9–16. doi: 10.1007/978-3-319-67159-8_2

Zhao, F., Chen, Z., Rekik, I., Lee, S. W., and Shen, D. (2020). Diagnosis of autism spectrum disorder using central-moment features from low- and high-order dynamic resting-state functional connectivity networks. *Front. Neurosci.* 14:258. doi: 10.3389/fnins.2020.00258

Zhao, F., Qiao, L., Shi, F., Yap, P. T., and Shen, D. (2017). Feature fusion via hierarchical supervised local CCA for diagnosis of autism spectrum disorder. *Brain Imaging Behav.* 11, 1050–1060. doi: 10.1007/s11682-016-9587-5

Zhao, F., Zhang, H., Rekik, I., An, Z., and Shen, D. (2018). Diagnosis of autism spectrum disorders using multi-level high-order functional networks derived from resting-state functional MRI. *Front. Hum. Neurosci.* 12:184.

Zhou, Y., Zhang, L., Teng, S., Qiao, L., and Shen, D. (2018). Improving sparsity and modularity of high-order functional connectivity networks for MCI and ASD identification. *Front. Neurosci.* 12:959. doi: 10.3389/fnins.2018.00959

Conflict of Interest: DS was a consultant of Shanghai United Imaging Intelligence Co., Ltd. The commercial affiliation did not play a role in the study design, data collection and analysis, decision to publish, or preparation of the 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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 25 March 2021
doi: 10.3389/fnins.2021.655599

Motor Imagery EEG Signal Recognition Using Deep Convolution Neural Network

*Xiongliang Xiao[1] and Yuee Fang[2]**

[1] College of Electronic Science and Engineering, Hunan University of Information Technology, Changsha, China, [2] College of Electric Power Engineering, Hunan Polytechnic of Water Resources and Electric Power, Changsha, China

Brain computer interaction (BCI) based on EEG can help patients with limb dyskinesia to carry out daily life and rehabilitation training. However, due to the low signal-to-noise ratio and large individual differences, EEG feature extraction and classification have the problems of low accuracy and efficiency. To solve this problem, this paper proposes a recognition method of motor imagery EEG signal based on deep convolution network. This method firstly aims at the problem of low quality of EEG signal characteristic data, and uses short-time Fourier transform (STFT) and continuous Morlet wavelet transform (CMWT) to preprocess the collected experimental data sets based on time series characteristics. So as to obtain EEG signals that are distinct and have time-frequency characteristics. And based on the improved CNN network model to efficiently recognize EEG signals, to achieve high-quality EEG feature extraction and classification. Further improve the quality of EEG signal feature acquisition, and ensure the high accuracy and precision of EEG signal recognition. Finally, the proposed method is validated based on the BCI competiton dataset and laboratory measured data. Experimental results show that the accuracy of this method for EEG signal recognition is 0.9324, the precision is 0.9653, and the AUC is 0.9464. It shows good practicality and applicability.

Keywords: EEG signal, motor imagination, deep convolutional neural network, short time fourier transform, continuous morlet wavelet transform, BCI classifier, CSP algorithm

OPEN ACCESS

Edited by:
Kaijian Xia,
Changshu No.1 People's Hospital, China

Reviewed by:
Jun Hu,
Wuhan University of Technology, China
Jin Yang,
Central South University, China

****Correspondence:***
Yuee Fang
zhangyusha82@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *19 January 2021*
Accepted: *03 March 2021*
Published: *25 March 2021*

Citation:
Xiao X and Fang Y (2021) Motor Imagery EEG Signal Recognition Using Deep Convolution Neural Network. Front. Neurosci. 15:655599. doi: 10.3389/fnins.2021.655599

INTRODUCTION

Brain-Computer Interface (BCI) is a communication control system established between the brain and external devices (computers or other electronic devices) through signals generated during brain activity (Gao, 2008; Mohamed et al., 2017; Wang et al., 2019). The system does not rely on muscles and nerves other than the brain, and establishes direct communication between the brain and the machine. It is a new and high-end human-computer interaction method.

The Motor Imagery Brain-Computer Interface (MI BCI) based on electroencephalogram (EEG) belongs to the category of spontaneous brain-computer interface (Abiri et al., 2018). The purpose of MI BCI is to accurately identify the user's physical movement intentions, which commonly include imagination of left and right hands, feet, and tongue movements. This is of great significance to the fields of medical rehabilitation, leisure and entertainment (Vasilyev et al., 2017; Minkyu et al., 2018). **Figure 1** is a structural diagram of a simple motor imagery recognition system. However, due to the non-stationary, non-linear, low signal-to-noise ratio and other characteristics of EEG

signals, there are still many problems to be solved in terms of preprocessing, feature extraction, and multi-mode classification (Yazici et al., 2019). As a result, there are fewer BCI systems that can be practically applied.

The EEG signal is a multi-channel signal, and there is no perfect theory for the role of each channel in classification. For EEG signal identification and classification, the traditional methods are mostly to manually select channels for experiments. In this process, it is possible to lose effective features or introduce unnecessary noise (Duan et al., 2020). At the same time, the EEG signal has its own low signal-to-noise ratio and other characteristics, plus interference from external environmental factors. It is also difficult to process its signal characteristics, which makes it difficult for its recognition accuracy to meet actual needs (Kim et al., 2018).

With the rise of big data and artificial intelligence technologies, deep learning algorithms have developed rapidly recently. Fruitful results have been achieved in the field of computer vision and speech recognition. Among them, the deep convolutional network represented by the CNN algorithm has low network complexity and strong feature extraction capabilities, which can well solve the problem of difficult feature extraction of EEG signals. Therefore, it is particularly urgent and feasible to recognize EEG signals based on deep convolutional networks.

RELATED RESEARCH

The significance of studying brain-computer interface technology is not limited to promoting the development of rehabilitation medicine. More importantly, it opens up a new way for people to obtain brain information, and at the same time enriches the content of brain cognitive science and neuroinformatics. It has huge research prospects, important theoretical value and practical significance. As the basic research in brain-computer interface, the research of EEG signal recognition is a multidisciplinary problem (Aznan et al., 2016; Xu et al., 2020). The key lies in how to extract and classify EEG signals accurately and effectively. At present, many scholars have carried out in-depth research on this.

Most existing feature extraction methods rely on human knowledge and experience. However, due to the limitations of human knowledge and experience, artificially designed features have certain limitations, and they cannot extract suitable features well, resulting in limited classification accuracy (Wang and Bezerianos, 2017). In addition, the process of finding suitable features usually requires some additional experiments, which will take a lot of time and energy. Traditional EEG signal feature recognition research mostly uses Short Term Fourier Transform (STFT) or Wavelet Transform (WT) to extract the time-frequency features of EEG signals (Tabar and Halici, 2017; Lee and Choi, 2019; You et al., 2020). However, these tasks are generally based on the extraction of time-frequency features in a fixed time period of EEG data or in the same frequency band. There are limitations in extracting features of EEG data in fixed time bands and frequency bands. Based on the extracted time-frequency features, some researchers select features for different experimental subjects to improve the classification accuracy on motor imaging tasks. Luo et al. (2016) first uses wavelet packet decomposition technology to extract time-frequency features. Then use the dynamic frequency feature selection (DFFS) to select the feature with the highest classification accuracy for each experimental object. Li et al. (2017) first selects the time period with the highest correlation between the event-related desynchronization (ERD)/event-related synchronization (ERS) phenomenon in the collected EEG signals. Then use WPD to extract the time-frequency characteristics of the EEG signal. Finally, the feature selection algorithm is used to select the feature with the highest classification accuracy. Saa and Cetin (2012) proposed the Filter Bank Common Spatial Pattern (FBCSP). After filtering the original EEG signal with a set of filters, the CSP method is used to extract features on each filtered frequency band. Finally, the feature selection algorithm is applied on the basis of the extracted features. The above work research has improved the accuracy of some motor imaging tasks to a certain extent. However, since it takes a lot of time to select the characteristics of the experimental data set for each experimental object, it is not universal.

In recent years, deep learning has been widely used in computer vision, speech recognition and recommendation systems, and has achieved great success. Because the deep learning method can automatically extract the input signal features, it avoids the limitations of manual design features. Therefore, some scholars apply deep learning algorithms to the classification of motor imagery EEG signals (Chu et al., 2018; He et al., 2020). Among the many deep learning algorithms, Convolutional Neural Network (CNN) has become the most popular method in the motor imagination EEG classification algorithm because of its excellent feature extraction capabilities. Literature (Lawhern et al., 2018) proposed a CNN structure that can be applied to a variety of popular brain-computer interface paradigms (including motor imagination, P300 visual evoked potentials, etc.). A higher classification accuracy rate than the FB-CSP method is obtained. Li (2020) proposed a moving image classification algorithm based on spatiotemporal features extracted by convolutional neural network. The temporal and spatial characteristics of the EEG signal are extracted by the longitudinal convolution kernel and the lateral convolution kernel, respectively. And built a five-layer neural network model to classify EEG signals. Wei et al. (2019) uses EEG emotion data set SEED for emotion recognition research. The abstract features of EEG samples are automatically extracted based on the convolutional neural network in deep learning, eliminating the need for manual feature selection and dimensionality reduction. And with the most advanced methods at present, a considerable accuracy rate has been achieved. Li et al. (2020) proposed an algorithm combining continuous wavelet transform and simplified convolutional neural network to improve the recognition rate of MI-EEG signals. The feasibility of the algorithm is verified by the BCI dataset.

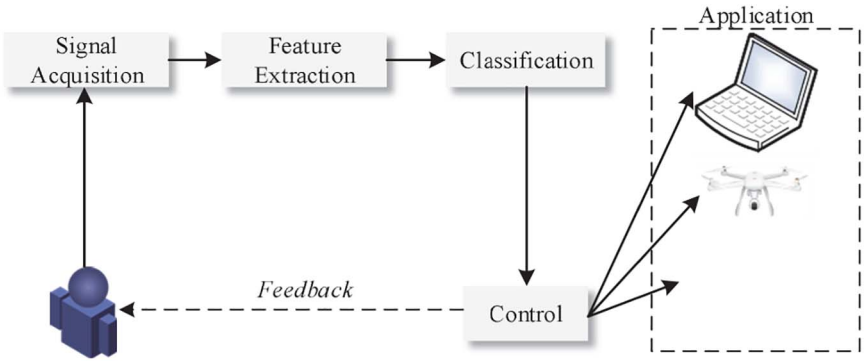


FIGURE 1 | Structure of motor imagery recognition system.

Drawing on the existing research work of EEG signal recognition, this paper proposes a motor imagery EEG signal recognition based on deep convolutional network. The main contributions are as follows:

1) Aiming at the difficult problem of EEG signal feature extraction, short-term Fourier transform is used to collect experimental data sets based on time series characteristics, so as to obtain EEG signals with time-frequency characteristics. And use continuous Morlet wavelet transform to further process the EEG signal difference of the data set. Provide high-quality data support for subsequent training and testing of deep convolutional network models.
2) It is oriented to the high accuracy and high precision requirements of EEG signal recognition for motor imagination. Based on the improved CNN network model, it realizes efficient recognition of EEG signals. Based on the advantages of the convolutional neural network's own network model, combined with the CSP algorithm, two-level feature extraction and classification are performed on the motor imagination EEG signal. Further improve the quality of EEG signal feature acquisition. Ensure the efficient performance of EEG recognition for ideal data sets and measured data sets.

The rest of this article is organized as follows. The third section introduces the data set used in this article and explains the corresponding data preprocessing methods. The fourth section introduces the main principles of EEG signal recognition based on the improved CNN network model. The fifth section carries out corresponding experimental simulation analysis on the feasibility and optimality of the proposed method. The sixth section is the conclusion and outlook.

DATASET PROCESSING

Sample Dataset

EEG signals are the distribution of potentials on the scalp produced by brain neuron activity, and are usually obtained by using an EEG device. The EEG data set used in this paper is BCI competiton data set.

The collection process of BCI competiton is described as follows: The subject wears an electrode cap and sits quietly in front of the computer, and imagines the movement of the left hand, right hand or right foot according to the prompts on the screen. Each subject performed a total of 280 motor imagination, 140 of which were left and right. The process of a single experiment lasting 7s is as follows:

(1) 1.5 s before the start of the experiment, a "++" prompt appeared on the screen to remind the subjects that they were about to perform the motor imagination task.
(2) At the beginning of the 1.5s of the experiment, the "+" disappeared, and L, R, F or arrows moving in different directions appeared in the center of the screen. According to the letters or moving arrows, the subjects imagined the movement of the left hand, right hand, and right foot. The process lasted 3.5 s.
(3) The arrows and letters on the screen disappear without any display. The subject can enter a relaxed state and rest for 2 s.

In the experiment, a 118-lead electrode cap was used to collect the EEG voltage on the scalp of the subject (Cheng et al., 2020). The acquisition frequency is 100 Hz. The EEG signals of all the above processes are collected by the system, so the EEG signals obtained are 5 matrices with 118 rows and 280*7*100 columns. Among them, the effective EEG signal of each subject performing the motor imaging task is a matrix of 118 rows and 280*3.5*100 columns.

During the experiment, the experimental paradigm shown in **Figure 2** was used (Graimann et al., 2003).

Data Preprocessing

When a person is performing limb motor imagination, a specific position in the motor sensory cortex of the brain will have regular potential changes (Pfurtscheller et al., 2005). When subjects perform unilateral limb motor imaging, the intensity of U rhythm (8–12 Hz) in the contralateral cortex of the brain decreases, and the intensity of Q rhythm (12–25 Hz) in the ipsilateral cortex increases, which is called event related

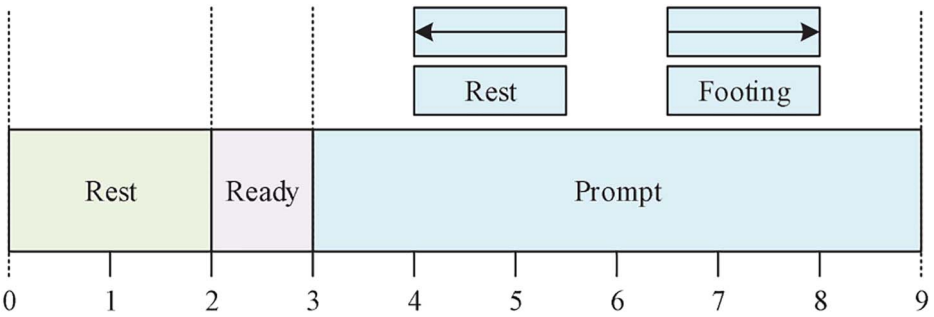


FIGURE 2 | MI experimental paradigm.

desynchronization (ERD) and event related synchronization (ERS) phenomenon (Mousavi and Sa, 2019). These two phenomena are important basis for distinguishing different types of EEG signals. Therefore, the time-frequency domain analysis method that combines the two is one of the most efficient analysis methods (Padfield et al., 2019).

Short-Time Fourier Transform

The short-time Fourier transform first divides the entire time series into several time segments of equal length. Then calculate the frequency spectrum information in each time segment by Fourier transform. Obtain the change of each frequency component with respect to time from the surface. The calculation formula is as follows:

$$S(f,k)=\sum_{n=0}^{N-1}s(n)\left[W(n-k)\,e^{\frac{f2\pi fn}{-N}}\right] \tag{1}$$

Where,$S(n)$ represents the time series of EEG signals.$W(n)$ represents window function.N represents the number of time points recorded.k represents the index of different time windows.f represents the frequency component in the signal.n represents time point. The length of the time window required to be divided in the formula is the same, which determines that the algorithm performs well when measuring high-frequency components. When measuring low-frequency components, it is often accompanied by distortion.

In order to effectively measure the change trend of the μ rhythm and β rhythm in the signal, this paper selects the time-frequency matrix obtained by the time window of 0.5s and the hamming window function (Soroosh and Mohammadi, 2018). Combine the time-frequency matrices on the two channels C3 and C4. A three-dimensional tensor with a size of 33*35*2 is obtained as the input of the subsequent convolutional neural network.

Continuous Morlet Wavelet Transform

The Morlet wavelet transform uses a wavelet of finite length and attenuation as the base to measure the intensity of each rate component in the signal over time. The formula is as follows:

$$W(a,b)=\int_{-\infty}^{\infty}x(t)\,\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \tag{2}$$

Where,$x(t)$ represents the signal sequence.$\psi(t)$ represents the wavelet basis.t represents the time point. The parameter a controls the scaling of the wavelet function. When a takes a value from small to large, the wavelet function gradually widens, so the low-frequency components can be better measured. And by adjusting the parameter b, the shift of the wavelet function is controlled to obtain the intensity information of each frequency band at different time domain positions. The calculation formulas of Morlet wavelet center time and time domain span are as follows:

$$\psi(x)=e^{-x^2}\cos\left(\pi\sqrt{\frac{2}{\ln 2}}x\right) \tag{3}$$

$$t_0=\frac{\int_{-\infty}^{\infty}t\,|\psi(t)|^2\,dt}{\int_{-\infty}^{\infty}|\psi(t)|^2\,dt} \tag{4}$$

$$\Delta t_\psi=\sqrt{\frac{\int_{-\infty}^{\infty}(t-t_0)^2\,|\psi(t)|^2\,dt}{\int_{-\infty}^{\infty}|\psi(t)|^2\,dt}} \tag{5}$$

The calculation formula for center frequency and bandwidth is as follows:

$$\omega_0=\sqrt{\frac{\int_{-\infty}^{\infty}\omega\,|\Psi(\omega)|^2\,d\omega}{\int_{-\infty}^{\infty}|\Psi(\omega)|^2\,d\omega}} \tag{6}$$

$$\Delta\omega_\psi=\sqrt{\frac{\int_{-\infty}^{\infty}(\omega-\omega_0)^2\,|\Psi(\omega)|^2\,d\omega}{\int_{-\infty}^{\infty}|\Psi(\omega)|^2\,d\omega}} \tag{7}$$

Where, $\Psi(\omega)$ is the frequency component information obtained after $\psi(t)$ undergoes Fourier transform. It can be known from the above formula that when the wavelet transform measures high frequency components, because the wavelet used is narrow, a smaller time domain span can be obtained, but the frequency domain span will be enlarged accordingly. Therefore, in the output time-frequency matrix, the resolution of the frequency dimension of the high frequency part is relatively low, and the low frequency part is just the opposite. Similarly, the C3 and C4 channel position information are integrated, and a sample matrix of size $35 \times 1152 \times 2$ is obtained as the input of the neural network.

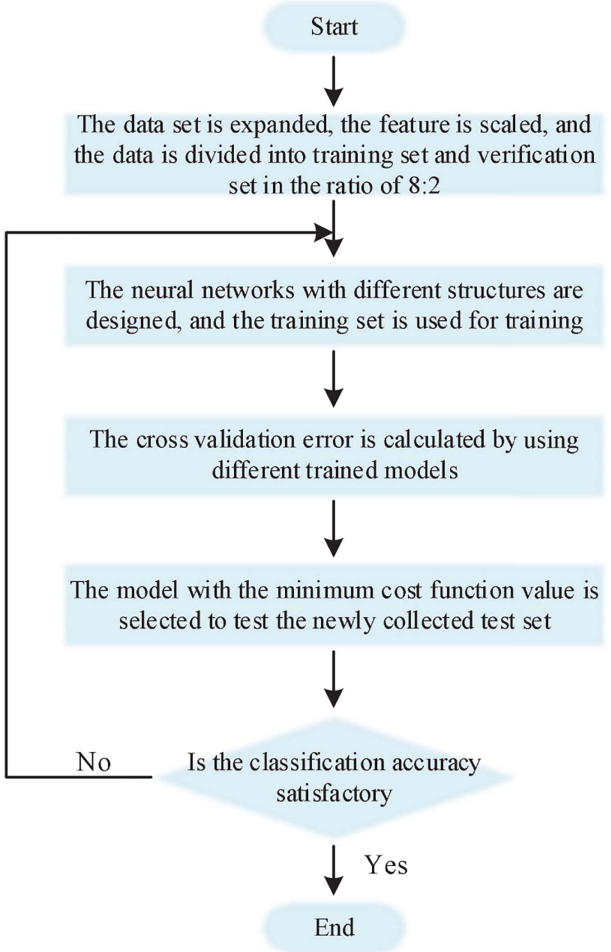


FIGURE 3 | Flow chart of BCI classifier design and training.

EEG SIGNAL CLASSIFICATION BASED ON IMPROVED CNN NETWORK MODEL

BCI Classifier Design and Training Process

The ability of deep learning algorithms to extract features is greatly improved compared with traditional algorithms. And generally, the more complex the network, the more sufficient features are extracted, and the better the result of the classifier. However, the advantages of the classification accuracy of deep learning algorithms are usually only reflected when the number of sample sets is large enough. And the more complex the network, the more parameters to be trained, the more training set samples will be needed.

It should be noted that the high complexity of the network model cannot be pursued blindly in the design of neural networks. We should balance the network structure and the number of sample sets, and design a preliminary neural network in advance (Anuse and Vyas, 2016; Ha and Jeong, 2019). Through the learning curve of the training set and the cross-validation set during the training process, it is judged whether the network is in an over-fitting or under-fitting state. Then debug the hyperparameters in the network according to the network status, and decide whether to increase the number of samples and how to modify the network structure. Until a network model with satisfactory classification effect is designed. **Figure 3** is a flowchart of the brain-computer interface classifier design used in this article.

Construction of Convolutional Neural Network

In this paper, the classic CNN structure is modified, and the input sample data is reduced from the traditional two-dimensional to one-dimensional. The reconstructed CNN is used to extract and classify the one-dimensional feature data obtained after the motion imagination EEG is processed by the CSP algorithm. After the one-dimensional feature data of EEG is subjected to the feature extraction process again through the convolutional layer, the fully connected layer and the Softmax classifier are used to output the classification results. The CNN structure of this paper is shown in **Figure 4**.

The improved CNN structure is mainly divided into 5 layers, the first layer is the input layer (I1). The input sample data size is 1 × N. Among them, N is the number of features obtained after the motion image EEG is processed by the CSP algorithm, $N = 4 \times m$; The second layer (C2) and the third layer (C3) are both convolutional layers for feature extraction of input sample data. The second layer (C2) has i2 convolution kernels, and the size of the convolution kernel is 1 × n2. The third layer (C3) has i3 convolution kernels with a size of 1 × n3. Due to the small length of the input sample data, the downsampling layer is omitted in this CNN; The fourth layer (F4) forms a single-layer perceptron together with the fifth layer (O5) in a fully connected way. After processing the output result of the third layer (C3), the classification result is output. The convolutional network is shown in Eq. 8.

$$CNN = [I1(1 \times N) - C2(i_2 \times 1 \times n_2)] - C3(i_3 \times 1 \times n_3) - F4 - O5] \quad (8)$$

It is assumed that the convolution kernel of each layer is represented by a matrix of size $[i_1^*, n_1]$, where i_1 represents the number of convolution kernels of the first convolution layer, and n_1 represents the length of the one-dimensional convolution kernel. When training starts, these convolution kernels are initialized to random values between [-0.3, +0.3]. Then the error value Error between the network predicted category Y_i and the sample actual category R_i is corrected through the error back propagation algorithm, where Error is as in Eq. 9. If $W_i^{l\phi}$ and $b_i^{l\phi}$ represent the weight and deviation of the i-th convolution kernel of the l-th convolutional layer, the feature map $H_i^{l\phi}$ is as in Eq. 10. The input and output of the j-th neuron in the fourth (F4) and fifth (O5) layers of the fully connected layer are $x^l(j)$ and $y^l(j)$

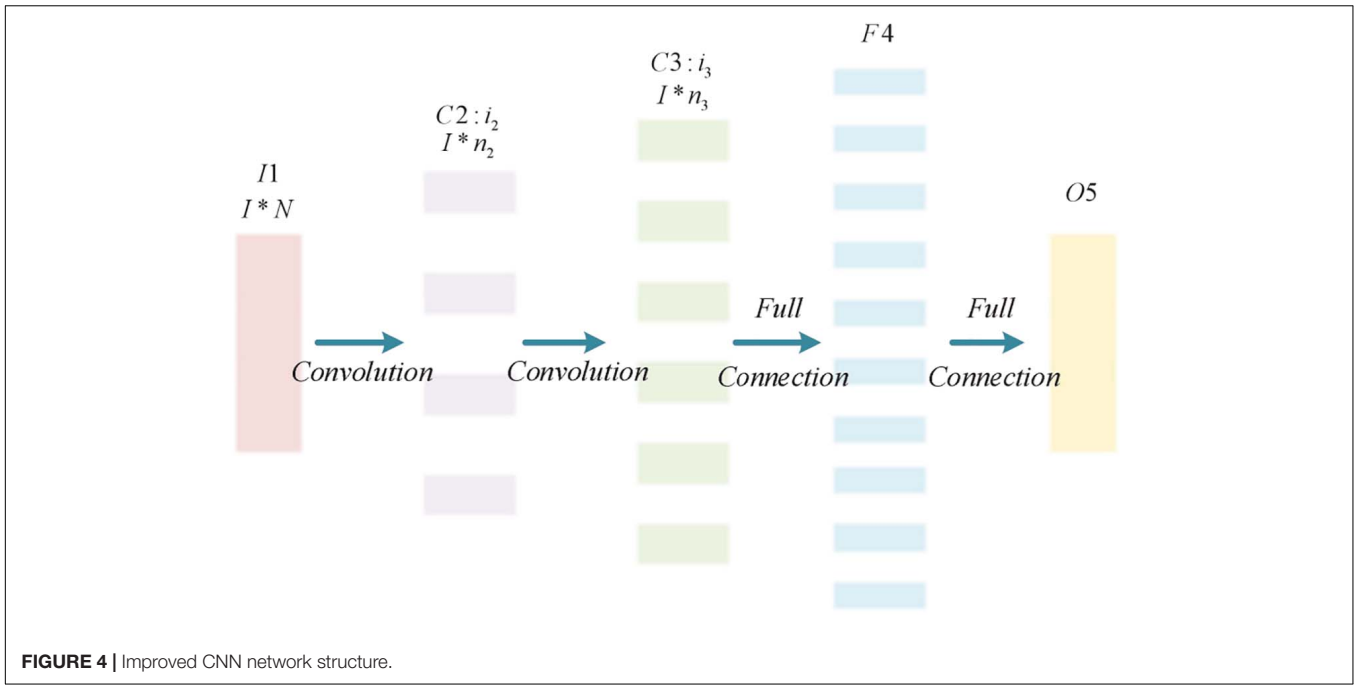


FIGURE 4 | Improved CNN network structure.

respectively. The hyperbolic tangent function is used as the neuron activation function, and the connection method is shown in Eq. 11.

$$Error = \sum_{i=1}^{n} (Y_i - R_i)^2 \tag{9}$$

$$H_i^{l\phi} = H_i^{(l-1)\phi} W_i^{l\phi} + b_i^{l\phi} \tag{10}$$

$$y^l(j) = \tanh[x^l(j) W^l + b^l(j)] \tag{11}$$

Network Training

In the training process of the improved CNN classifier in this paper, there are still a large number of hyperparameters that need to be manually set. The specific hyperparameters related to network training in this experiment will be set as follows.

(1) Batch size and epoch: The batch size is the number of samples used in a training process. The value of the number of epochs is equal to all training samples divided by the batch size. In the training process of deep learning, all training sets are usually not loaded into memory at once for iterative calculations. Because the total number of training sets is too large, it will cause problems such as low memory efficiency and slow training speed. Experimenters need to consider the server memory size, input sample size, network model complexity and other factors to choose an appropriate batchsize. Make the network only read batchsize training samples during each training process. Setting the Batchsize too small will cause the network to be difficult to converge and underfit. If the batchsize is set too large, it will result in reduced efficiency or memory overflow. In this experiment, the batchsize is 64 and the epoch is 150.

(2) Learning rate: Learning rate is a very important hyperparameter in network training. Whether it is set reasonably can directly affect the final classification accuracy of the network. If the learning rate is set too small, the error curve will fall too slowly, and the learning rate will be too large, which will cause the error to explode, and the network cannot correctly find the direction of the gradient drop. After many attempts and comprehensive considerations, this experiment dynamically changes the learning rate during the training process. Set the initial value of the learning rate to 0.02 and the end value to 0.0002. Decay in the form of math.exp (- /decay_ speed). Where i is the number of iterations, decay_speed is the decay speed, and the value is 1,000 in this paper.

(3) Initialization of weights and biases: In this experiment, the weight $w_{i,j}$ is initialized as normal random initialization, and the bias w_b is initialized as a constant matrix 0.1.

(4) Dropout:Overfitting often occurs in the training process of deep learning. Corresponding solutions include increasing the number of training set samples, adding a regularization function, and adopting methods such as Dropout. Since the convolutional neural network uses Relu as the activation function, the sparsity of the function makes the network self-regularized. Therefore, this experiment uses the Dropout method to set its value to 0.5, so that 50% of the hidden layer nodes do not work in the layer that uses Dropout during each training process. Thereby reducing the phenomenon of over-fitting and enhancing the expressive ability of the network model.

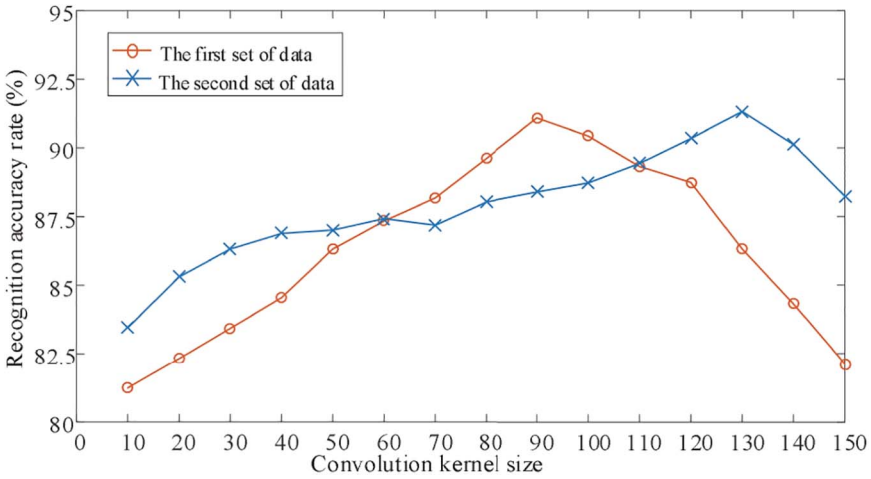


FIGURE 5 | Influence of convolution check on identification results.

After designing the CNN network model and assigning all the hyperparameters in the model, start training the classifier. Convolutional networks are trained using BP algorithm (Jia et al., 2019; Liu et al., 2020), and the process is divided into three steps: The forward propagation calculates the output value a_{ij}, the back propagation calculates the error δ_{ij}, the gradient of the weight w_{ij} is calculated and the weight is updated.

The output value $a_{d,i,j}$ of the forward propagation process of the convolutional layer is calculated as follows. Where$D$is the input depth and F is the size of the convolution kernel.$w_{d,m,n}$ represents the weight of the mth row and nth column of the d-th convolution kernel.$x_{d,i,j}$ represents the element in row i and column j of the d layer input.f is the Relu activation function.w_b is the bias of the convolution kernel.

$$a_{d,i,j} = f\left(\sum_{d=0}^{D-1}\sum_{m=0}^{F-1}\sum_{n=0}^{F-1} w_{d,m,n}x_{d,i+m,j+n} + w_b\right) \quad (12)$$

Because the CNN network uses the connection method of the local receptive field and the down-sampling operation such as pooling processing, the calculation method of the error term of the model is very different from the traditional fully connected network. The derivation process of the $l-1$-th layer error term δ^{l-1} in the back propagation process of the CNN convolutional layer is shown in the following Eqs 13–16, where E_d represents the error function. The calculation of the error term of the CNN pooling layer depends on the specific pooling method. The error term of the maximum pooling is transferred from the next layer to the position corresponding to the upper layer $\delta_{i,j}^{l-1}$ intact, and the error term of the remaining positions is set to 0. The average pooling error term δ^{l-1} is evenly distributed from the next layer δ^{l} to each neuron corresponding to the previous layer.

$$net^l = conv\left(W^l, a^{l-1}\right) + w_b \quad (13)$$

$$a_{i,j}^{l-1} = f^{l-1}\left(net_{i,j}^{l-1}\right) \quad (14)$$

$$\begin{aligned}\delta_{i,j}^{l-1} &= \frac{\partial E_d}{\partial net_{i,j}^{l-1}} = \frac{\partial E_d}{\partial a_{i,j}^{l-1}}\frac{\partial a_{i,j}^{l-1}}{\partial net_{i,j}^{l-1}} \\ &= \sum_m \sum_n w_{m,n}^l \delta_{i+m,j+n}^l f'\left(net^{l-1}\right)\end{aligned} \quad (15)$$

$$\delta^{l-1} = \sum_{d=0}^{D}\sum_m\sum_n w_{m,n}^l \delta_{i+m,j+n}^l f'\left(net^{l-1}\right) \quad (16)$$

The last step of training is to calculate the gradient and update the weight accordingly. The pooling layer does not introduce the parameters to be learned, only the gradient of the weight and bias of the convolution layer needs to be calculated. Due to weight sharing, each weight $w_{i,j}$ has an effect on each $net_{i,j}^l$, so its calculation formula is shown in the following Eqs 17 and 18. Finally, the network uses the gradient descent method to update the weights, as shown in Eq. 19. In the equation, η is the learning rate. Through repeated iterations of a large number of samples, the error function value is continuously reduced every time the weight is updated. Training of the CNN model.

$$\begin{aligned}\frac{\partial E_d}{\partial w_{i,j}} &= \sum_m \sum_n \frac{\partial E_d}{\partial net_{m,n}^l}\frac{\partial net_{m,n}^l}{\partial w_{i,j}} \\ &= \sum_m \sum_n \delta_{m,n} a_{i+m,j+n}^{l-1}\end{aligned} \quad (17)$$

$$\frac{\partial E_d}{\partial w_b} = \sum_m\sum_n \delta_{m,n}^l \quad (18)$$

$$w_{i,j} = w_{i,j} - \eta\frac{\partial E_d}{\partial w_{i,j}} \quad (19)$$

EXPERIMENT AND ANALYSIS

In order to verify the feasibility and practicability of the method mentioned above, the experimental simulation

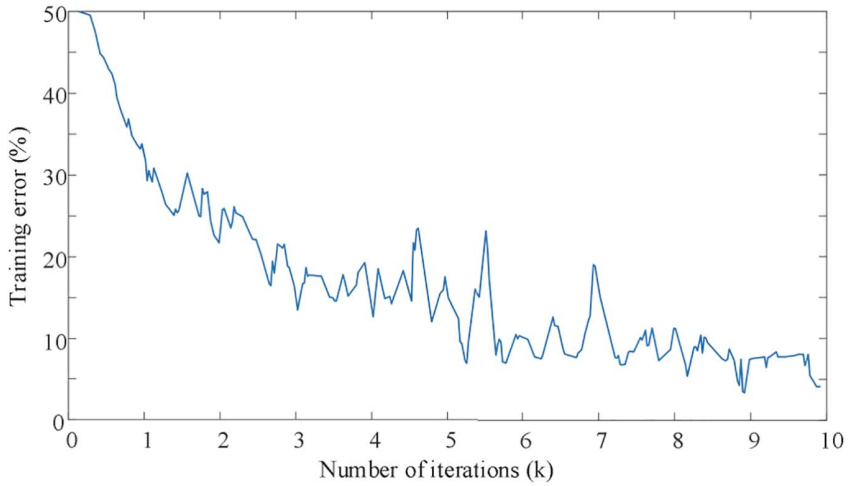
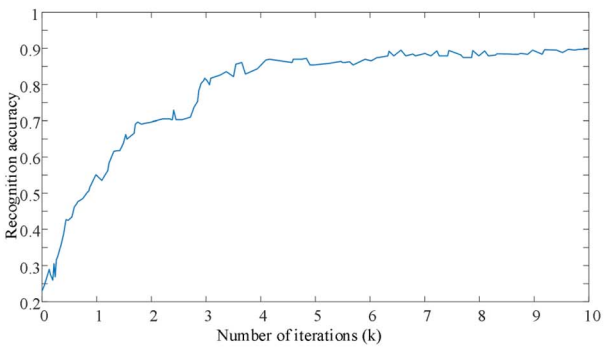


FIGURE 6 | The curve of training error with iteration times.



A Classification accuracy of training set



B Classification accuracy of test set

FIGURE 7 | The curve of recognition accuracy with the number of iterations.

robot hardware environment is: Lenovo Tinkpad E14, AMD Ryzen 7 4700U 8-core processor, 16GB memory, integrated graphics; The software environment is: operating system Chinese Windows 10, English version software Microsoft Visual Studio 2012.

This article uses the Caffe deep learning framework to implement model training and testing on the data set used in this article. The data samples in the experimental data set are divided into training set and cross-validation set at a ratio of 8:2, and 6,880 training set samples and 1,720 validation set samples are obtained. The deep convolutional network parameters are set according to the corresponding network model described in the section "Network Training."

Network Model Optimization and Analysis

The influence of Convolution Kernel Size on the Accuracy of EEG Classification

For the CNN network, determining the appropriate size of the convolution kernel is of great significance to the improvement of feature extraction and recognition accuracy. Therefore, for the classification of motor imagery EEG signals, this article first studied the influence of different convolution kernel scales on the identification of experimental data sets. The experimental results are shown in **Figure 5**.

It can be seen from **Figure 5**, where **Figure 5** shows the changes in the classification accuracy of the motor imagination EEG signals of experimental objects as the size of the convolution kernel changes. It can be seen from **Figure 5** that as the size of the convolution kernel increases, the classification accuracy of the EEG signal of the BBCI competiton dataset has just begun to gradually increase. Then gradually decrease, the optimal convolution kernel size is 1×115. The EEG signal classification accuracy rate of the measured data set increases first and then decreases as the size of the convolution kernel increases, but the optimal convolution kernel size is 1×55. This shows that for different experimental subjects, convolution kernels of different sizes are needed to extract the

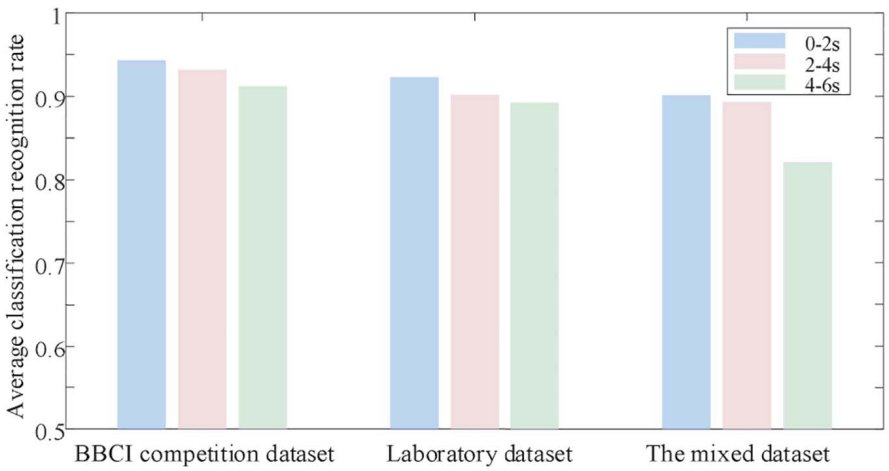


FIGURE 8 | Classification results analysis based on temporal sequence.

TABLE 1 | Evaluation index of control models.

Method	Accuracy	Discrimination	Sensitivity	Specificity	AUC
The proposed method	0.9324	0.9653	0.8682	0.9243	0.9464
Lawhern et al. (2018)	0.9135	0.9211	0.8732	0.8932	0.9132
Li (2020)	0.8932	0.8976	0.8321	0.8589	0.8843
Wei et al. (2019)	0.8821	0.8832	0.7932	0.7932	0.8591

most suitable features for motor imagination brain electrical signal classification.

The Training Process of the Subject CNN Classifier

In order to further prove the feasibility and superiority of the method proposed in this paper for accurate identification of EEG signals, the experiment intends to analyze the training error and training accuracy of EEG signal identification under different iteration times. The experimental results are shown in **Figures 6**, **7**.

As shown in **Figure 6**, it can be seen that the EEG signal data set is oriented to mixed motion imagination. Even though the training set data volume has high-dimensional and large-scale characteristics, the proposed improved CNN network model also shows excellent characteristics for the convergence speed of the recognition process. When the number of iterations of the improved CNN network reaches 8,000 times, the training error remains below 10%. At the same time, it can also be seen from **Figure 7A** that the improved CNN network has basically achieved full signal recognition after 9,000 iterations for the motion imaging training set data signal recognition accuracy used in this article. The identification of the sample test data set in **Figure 7B** also shows efficient convergence characteristics. That is to say, in 8,000 iterations, the recognition accuracy of EEG signals of motor imagination can achieve an effective recognition of more than 95%. Therefore, it is confirmed that the EEG signal recognition method proposed in this paper has the advantages of fast and efficient convergence characteristics.

Identification and Analysis of Mixed Datasets

Analysis of Classification Results Based on Time Series

The experiment divides the EEG experimental data of motor imagination into time periods, each with a duration of 2 s as input data. **Figure 8** shows the average classification recognition rate of the tested users in the 3 time periods from 0 to 6 s.

It can be seen from **Figure 8** that the average classification and recognition rate of EEG data in the first 2 s (0th–2nd s) is the highest. The average classification and recognition rate of EEG data in the last 4 s (2–6 s) is low. Explain that at the beginning of the experiment, the subject users focused on the motor imagination experiment. However, as time goes by, the concentration of the tested users decreases, which affects the quality of EEG data and ultimately leads to a decrease in the recognition rate. Therefore, according to the above analysis results, the original input data dimension is selected as 32×64, convolution and pooling 3-layer network, and the EEG data from 0 to 2 s after the start of motor imagination is selected for classification prediction.

Performance Comparison With Several Comparison Algorithms on Public Datasets

In order to verify the superiority of the deep convolutional network proposed in this paper in EEG signal recognition, this paper reproduced three other analysis methods based on the mixed motor imagery data set [Literature (Lawhern et al., 2018;

Wei et al., 2019; Li, 2020)]. And get its evaluation index on the test set as shown in **Table 1**.

It can be seen from **Table 1** that the accuracy index of the method proposed in this paper is 0.9324, which is higher than the literature (Lawhern et al., 2018) 0.0189, literature (Li, 2020) 0.0392, and literature (Wei et al., 2019) 0.0503. In terms of accuracy indicators, the improved CNN network is 0.0442, 0.0677, and 0.0821 higher than the comparison algorithm. As for the sensitivity index, the method proposed in this paper is not outstanding compared with the comparison algorithm, which is 0.005 lower than that in the literature (Lawhern et al., 2018). The fundamental reason is that the improved deep convolutional network has a deeper network depth, so that EEG signal recognition can guarantee higher accuracy during training and testing. But through the continuous learning and training of the deep network, this reduces the sensitivity of the network model to a certain extent. For the specificity index and AUC index, the methods proposed in this paper are 0.9243 and 0.9464, respectively, which are 0.0311 and 0.0332 higher than those in the literature (Lawhern et al., 2018).

CONCLUSION

Motor imaging EEG signal recognition is an important and challenging research problem in human-computer interaction. Facing the accuracy and precision requirements of emotion recognition, this paper combines neural network and proposes a motor imagery EEG signal recognition method based on deep convolutional network. This method first uses short-time Fourier transform and continuous Morlet wavelet transform to preprocess the collected experimental data sets, so as to provide high-quality EEG signals for subsequent network models. Then, based on the improved CNN network model, the processed EEG signals are efficiently identified. Improve the quality of EEG signal feature acquisition and ensure the high accuracy and precision of EEG signal recognition. According to the analysis of the experimental results, the proposed method has an accuracy of 0.9324, an accuracy of 0.9653, and an AUC of 0.9464 for EEG signal recognition, and it has a good EEG signal recognition performance. The focus of future research will be to explore the platformization of the proposed method and strive to realize the commercialization of the proposed method. The focus of future research will be to explore the platformization of the proposed method and strive to realize the commercialization of the proposed method.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

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.

AUTHOR CONTRIBUTIONS

YF was proposed the main idea of this manuscript and completed the writing guidance, English polish and funding project. XX completed the algorithm design and experimental environment construction and completed the writing of the article. Both authors completed the experimental verification.

FUNDING

This work was supported by the Key projects of educational science research in Hunan Province (No. 20A347).

REFERENCES

Abiri, R., Borhani, S., Sellers, E. W., Yang, J., and Zhao, X. (2018). A comprehensive review of EEG-based brain-computer interface paradigms. *J. Neural Eng.* 16:011001. doi: 10.1088/1741-2552/aaf12e

Anuse, A., and Vyas, V. (2016). A novel training algorithm for convolutional neural network. *Complex Intell. Syst.* 2, 221–234. doi: 10.1007/s40747-016-0024-6

Aznan, N. K. N., Huh, K. M., and Yang, Y. M. (2016). EEG-based motor imagery classification in BCI system by using unscented Kalman filter. *Int. J. Inform. Commun. Technol.* 9, 492–501. doi: 10.1504/ijict.2016.10000493

Cheng, S. W., Chen, Y. J., and Xu, J. R. (2020). An eye movement direction classification method based on EEG. *Comput. sci.* 47, 118–124.

Chu, Y., Zhao, X., Zou, Y., Xu, W. L., Han, J. D., and Zhao, Y. W. (2018). A decoding scheme for incomplete motor imagery EEG with deep belief network. *Front. Neurosci.* 12:680. doi: 10.3389/fnins.2018.00680

Duan, L., Li, J., Ji, H., Pang, Z., and Zhuang, J. (2020). Zero-shot learning for EEG classification in motor imagery-based BCI system. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2411–2419. doi: 10.1109/tnsre.2020.3027004

Gao, S. (2008). Oscillatory EEG based brain-computer interface. *Chin. Bull. Life Sci.* 20, 722–724.

Graimann, B., Huggins, J. E., Schlogl, A., Levine, S. P., and Pfurtscheller, G. (2003). Detection of movement-related patterns in ongoing single-channel electrocorticogram. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 276–281. doi: 10.1109/tnsre.2003.816863

Ha, K. W., and Jeong, J. W. (2019). Motor imagery EEG classification using capsule networks. *Sensors* 19:2854. doi: 10.3390/s19132854

He, Q., Shao, D., and Wang, Y. (2020). Analysis and intention recognition of motor imagery EEG signals based on multi-feature convolutional neural network. *Chin. J. Sci. Instrum.* 41, 138–146.

Jia, H., Wang, S., Zheng, D., Qu, X., and Fan, S. (2019). Comparative study of motor imagery classification based on BP-NN and SVM. *J. Eng.* 2019, 8646–8649. doi: 10.1049/joe.2018.9075

Kim, C. S., Sun, J. W., Liu, D., Wang, Q. S., and Paek, S. (2018). An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI. *Med. Biol. Eng. Comput* 56, 1645–1658. doi: 10.1007/s11517-017-1761-4

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* 15:056013. doi: 10.1088/1741-2552/aace8c

Lee, H. K., and Choi, Y. S. (2019). Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface. *Entropy* 21:1199. doi: 10.3390/e21121199

Li, F., He, F., Wang, F., Zhang, D. Y., Xia, Y., and Li, X. Y. (2020). A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning. *Appl. Sci.* 10:1605. doi: 10.3390/app10051605

Li, M. A., Zhu, W., Liu, H. N., and Yang, J. F. (2017). Adaptive feature extraction of motor imagery EEG with optimal wavelet packets and SE-Isomap. *Appl. Sci.* 7:390. doi: 10.3390/app7040390

Li, X. L. (2020). Motor imagery-based EEG signals classification by combining temporal and spatial deep characteristics. *Int. J. Intell. Comput. Cybern.* 13, 437–453. doi: 10.1108/ijicc-07-2020-0077

Liu, F., Zhang, J., and Liu, T. (2020). A PSO-algorithm-based consensus model with the application to large-scale group decision-making. *Complex Intell. Syst.* 6, 287–298. doi: 10.1007/s40747-020-00144-5

Luo, J., Feng, Z., Zhang, J., and Lu, N. (2016). Dynamic frequency feature selection based approach for classification of motor imageries. *Comput. Biol. Med.* 75, 45–53. doi: 10.1016/j.compbiomed.2016.03.004

Minkyu, A., Hohyun, C., Sangtae, A., and Jun, S. C. (2018). User's self-prediction of performance in motor imagery brain–computer interface. *Front. Hum. Neurosci.* 12:59. doi: 10.3389/fnhum.2018.00059

Mohamed, S., Haggag, S., Nahavandi, S., and Haggag, O. (2017). Towards automated quality assessment measure for EEG signals. *Neurocomputing* 237, 281–290. doi: 10.1016/j.neucom.2017.01.002

Mousavi, M., and Sa, V. R. D. (2019). Spatio-temporal analysis of error-related brain activity in active and passive brain–computer interfaces. *Brain Comput. Interfaces* 6, 118–127. doi: 10.1080/2326263x.2019.1671040

Padfield, N., Zabalza, J., Zhao, H. M., Masero, V., and Ren, J. C. (2019). EEG-based brain-computer interfaces using motor-imagery:techniques and challenges. *Sensors* 19:1423. doi: 10.3390/s19061423

Pfurtscheller, G., Neuper, C., Brunner, C., and Silvac, F. L. (2005). Betarebound after different types of motor imagery in man. *Neurosci. Lett.* 378, 156–159. doi: 10.1016/j.neulet.2004.12.034

Saa, J. F. D., and Cetin, M. (2012). A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data. *J. Neural Eng.* 9:026020. doi: 10.1088/1741-2560/9/2/026020

Soroosh, S., and Mohammadi, A. (2018). Bayesian optimized spectral filters coupled with ternary ECOC for single trial EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2249–2259. doi: 10.1109/tnsre.2018.2877987

Tabar, Y. R., and Halici, U. (2017). A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* 14:016003. doi: 10.1088/1741-2560/14/1/016003

Vasilyev, A., Liburkina, S., Yakovlev, L., Perepelkina, O., and Kaplan, A. (2017). Assessing motor imagery in brain-computer interface training: Psychological and neurophysiological correlates. *Neuropsychologia* 97, 56–65. doi: 10.1016/j.neuropsychologia.2017.02.005

Wang, H., and Bezerianos, A. (2017). Brain-controlled wheelchair controlled by sustained and brief motor imagery BCIs. *Electron. Lett.* 53, 1178–1180. doi: 10.1049/el.2017.1637

Wang, Y., Nakanishi, M., and Zhang, D. (2019). EEG-based brain-computer interfaces. *Adv. Exp. Med. Biol.* 1101, 41–65.

Wei, C., Chen, L., and Zhang, A. (2019). EEG emotion recognition based on integrated convolutional neural network. *J. East China Uni. Sci. Technol. (Natural Science Edition)* 45, 614–622.

Xu, J., Zheng, H., Wang, J., Li, D., and Fang, X. (2020). Recognition of EEG Signal Motor Imagery Intention Based on Deep Multi-View Feature Learning. *Sensors* 20, 3496. doi: 10.3390/s20123496

Yazici, M., Ulutas, M., and Okuyan, M. (2019). A comprehensive sLORETA study on the contribution of cortical somatomotor regions to motor imagery. *Brain Sci.* 9:372. doi: 10.3390/brainsci9120372

You, Y., Chen, W., and Zhang, T. (2020). Motor imagery EEG classification based on flexible analytic wavelet transform. *Biomed. Signal Process. Control* 62:102069. doi: 10.1016/j.bspc.2020.102069

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 25 March 2021
doi: 10.3389/fnins.2021.662674

A Novel Brain MRI Image Segmentation Method Using an Improved Multi-View Fuzzy *c*-Means Clustering Algorithm

Lei Hua[1], Yi Gu[1], Xiaoqing Gu[2]*, Jing Xue[3] and Tongguang Ni[2]*

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China, [2] School of Computer and Artificial Intelligence, Changzhou University, Changzhou, China, [3] Department of Nephrology, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China

Background: The brain magnetic resonance imaging (MRI) image segmentation method mainly refers to the division of brain tissue, which can be divided into tissue parts such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). The segmentation results can provide a basis for medical image registration, 3D reconstruction, and visualization. Generally, MRI images have defects such as partial volume effects, uneven grayscale, and noise. Therefore, in practical applications, the segmentation of brain MRI images has difficulty obtaining high accuracy.

Materials and Methods: The fuzzy clustering algorithm establishes the expression of the uncertainty of the sample category and can describe the ambiguity brought by the partial volume effect to the brain MRI image, so it is very suitable for brain MRI image segmentation (B-MRI-IS). The classic fuzzy *c*-means (FCM) algorithm is extremely sensitive to noise and offset fields. If the algorithm is used directly to segment the brain MRI image, the ideal segmentation result cannot be obtained. Accordingly, considering the defects of MRI medical images, this study uses an improved multiview FCM clustering algorithm (IMV-FCM) to improve the algorithm's segmentation accuracy of brain images. IMV-FCM uses a view weight adaptive learning mechanism so that each view obtains the optimal weight according to its cluster contribution. The final division result is obtained through the view ensemble method. Under the view weight adaptive learning mechanism, the coordination between various views is more flexible, and each view can be adaptively learned to achieve better clustering effects.

Results: The segmentation results of a large number of brain MRI images show that IMV-FCM has better segmentation performance and can accurately segment brain tissue. Compared with several related clustering algorithms, the IMV-FCM algorithm has better adaptability and better clustering performance.

Keywords: brain magnetic resonance imaging, multi-view learning, fuzzy clustering, adaptive learning, image segmentation

OPEN ACCESS

Edited by:
Mohammad Khosravi,
Persian Gulf University, Iran

Reviewed by:
Yuexin Li,
Hubei University, China
Hongru Zhao,
Soochow University, China

****Correspondence:***
Yi Gu
8202101437@jiangnan.edu.cn
Xiaoqing Gu
guxq@cczu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *01 February 2021*
Accepted: *22 February 2021*
Published: *25 March 2021*

Citation:
Hua L, Gu Y, Gu X, Xue J and Ni T (2021) A Novel Brain MRI Image Segmentation Method Using an Improved Multi-View Fuzzy c-Means Clustering Algorithm. Front. Neurosci. 15:662674. doi: 10.3389/fnins.2021.662674

INTRODUCTION

The brain is the human body's nerve center, controlling people's thinking, memory, speech, and movement, and plays a role in regulating human organs. When the brain is healthy, it can work quickly and efficiently. However, when something goes wrong, the result can be devastating. In recent years, the problem of brain diseases has become increasingly prominent due to comprehensive factors such as high pressure in people's lives, fast-paced activities, extreme tension in thoughts and emotions, frequent accidents, and serious aging in the population, which continue to threaten people's health. In clinical medicine, doctors usually use MRI technology to diagnose brain diseases. MRI is a very advanced medical imaging technology. It visualizes the structure and function of the human body through radiology and is particularly suitable for brain tissue research. By using this technology, high soft tissue contrast can be obtained, and it has the advantages of noninvasiveness, nonradiation, and high-precision spatial resolution (Ji et al., 2011; Jiang et al., 2012; Thanh and Wu, 2013).

Neuroscience researchers need to segment brain MRI images to quantitatively study brain diseases. B-MRI-IS marks each pixel in the image as the corresponding brain tissue anatomical structure, such as the thalamus, hippocampus, and ventricles. Then, the segmented tissue size, shape, location, and other characteristics are used to evaluate and formulate medical plans. Experiments have shown that the abnormal shape or volume of certain anatomical regions of the brain is related to brain diseases, such as Alzheimer's disease and Parkinson's disease. The quality of B-MRI-IS determines the reliability of researchers' assessment of brain diseases. The segmentation quality is determined by the segmentation method, so the segmentation method is very important in the entire diagnosis process. Usually, accurate segmentation results of brain tissue are obtained by manual segmentation by experienced brain experts. However, when faced with a large number of datasets, manual segmentation methods become quite expensive, time consuming, and impractical. Moreover, due to differences in experience and knowledge among experts, the segmentation results are not uniform. Therefore, the use of more accurate B-MRI-IS algorithms to quantitatively analyze the volume shape of each tissue in brain MRI images has become the focus of medical image research.

Traditional image segmentation methods include threshold-based methods, clustering-based methods, region-based methods, edge-based methods, and graph theory-based methods (Gu et al., 2019; Kumar et al., 2020). Considering the problems in the segmentation process, scholars have proposed their solutions. Among them, fuzzy clustering algorithms are widely used. The literature (Bezdek, 1981) proposed the FCM algorithm for the first time, introducing the concept of ambiguity into the clustering method (Li et al., 2014), and proposed an optimized bias field estimation and tissue segmentation method. Through the simultaneous iteration of bias field estimation and tissue segmentation, the final segmentation accuracy is improved, and the algorithm is very robust to initialization. The literature (Chuang et al., 2006) proposed the FCM algorithm based on spatial relationships. The algorithm integrates the spatial information of the sum of membership degrees of all points in the window into the traditional FCM. Experiments show that the improved algorithm has certain robustness to noise. Noise interference has always been one of the important interference factors in medical image segmentation. Due to the complexity of the human brain tissue structure and the imaging mechanism of MRI technology, MRI images will show phenomena such as strong noise, false images, and weak boundaries. For example, traditional active contour models often only use target edge information, which leads to premature stopping when processing strongly noisy images, and weak boundary leakage occurs when the boundary is blurred. Many studies use grayscale information and spatial information to suppress noise interference. The literature (Ahmed et al., 2002) introduces local spatial terms into the objective function of FCM to make the algorithm perform better. The literature (Chen and Zhang, 2004) replaces the traditional Euclidean distance with the core distance to make the algorithm more robust. The literature (Krinidis and Chatzis, 2010) proposed a new blur factor and introduced the kernel distance, which is more robust to noise. The literature (Gong et al., 2013) proposed a kernel-based adaptive regularization FCM algorithm. This method introduces adaptive parameters and uses mean filtering, median filtering, and custom filtering of images. There is no need to calculate parameters for each iteration, which greatly saves time and cost. Additionally, the robustness of the algorithm to noise is improved. The literature (Elazab et al., 2015) improves the quality and compression rate of the reconstructed image by compressing the image in blocks.

Figure 1 shows the motivation for this research. To reduce the influence of noise in the segmentation process, the accuracy of B-MRI-IS is improved. This article uses an improved multiview FCM algorithm. The work in this paper is summarized as follows:

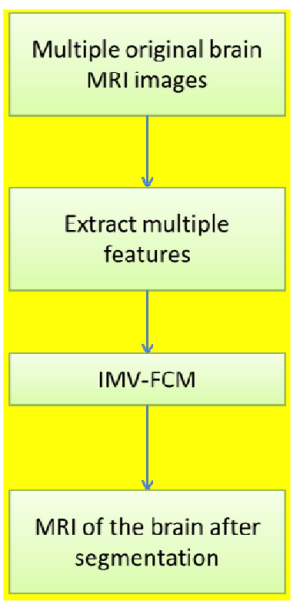


FIGURE 1 | The motivation for this research.

1. The B-MRI-IS process is proposed. First, the original brain MRI image is preprocessed, such as denoising. Second, the histogram of oriented gradient (HOG) feature, entropy feature, gradient feature, and contrast feature of the original image are extracted. Each feature data point is used as view data to construct multiview data (MVD). Third, the constructed MVD is input into the clustering model to obtain the final clustering result. The clustering result is the image segmentation result.
2. Based on multiview FCM, the IMV-FCM algorithm is proposed. IMV-FCM uses a view weight adaptive learning mechanism so that each view obtains the optimal weight according to its cluster contribution. The final division result is obtained through the view ensemble method. Under the view weight adaptive learning mechanism, the coordination between various views is more flexible, and each view can be adaptively learned to achieve better clustering effects.
3. To demonstrate the effectiveness of the clustering model used in this paper, single-view clustering and multiview clustering algorithms are used for experimental comparison. To demonstrate the antinoise of the algorithm, after adding noise to the experimental image, each algorithm is used for clustering comparison. The results show that the algorithm used in this paper is the most robust to noise compared with other comparison algorithms and can accurately segment brain tissue.

BACKGROUNDS

MRI Technology

In medicine, human brain tissue includes the cerebrum, cerebellum, and brainstem, each of which has relatively independent characteristics. At the same time, the skull structure can be divided into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) (Wang et al., 2015). GM contains WM, and CSF fills the crooked groove formed by the GM folds. The brain tissue images show that the gray levels of various brain tissues change slowly and are not constant. Coupled with the influence of various noises, it is easy to cause the gray values to cross and overlap the internal brain tissues. The only correspondence is that one gray value may correspond to several brain internal tissues. In addition, internal brain images with different imaging morphologies also provide different information, and they appear as different gray levels on the images. According to the image gray level, it can be divided into four types, namely, white, off-white, gray, and black. **Table 1** shows the characteristics of different gray levels.

The proton density ρ, longitudinal relaxation time T1, and transverse relaxation time T2 of brain tissue constitute the high-dimensional biophysical property space of the brain. The T1 and T2 parameters of normal and pathological tissues in different organs of the human body are fixed, and there are certain differences between them. Additionally, different tissues have different densities, and the differences in these properties between brain tissues are the biological basis of the solvability of brain tissue segmentation. **Table 2** shows the relevant attributes of the main brain tissues in MRI images.

TABLE 1 | Image grayscale characteristics of adult brain tissue.

Organization	T1-weighted image	Pd weighted image	T2 weighted image
WM	Off-white	Gray	Black gray
GM	Gray	Off-white	Off-white
CSF	Black	Gray	White

Clustering Algorithm

These attributes are determined by factors such as the composition of the brain tissue and the local microstructure. Cluster analysis aims to make reasonable classifications according to the characteristics of the sample and divide the data points with similar characteristics into one category. The idea of the cluster analysis algorithm comes from taxonomy. As MRI images contain increasing information, the requirements for classification are increasing. People have introduced multivariate analysis into taxonomy, forming cluster analysis. The goal of cluster analysis is to collect data to classify when the data points are similar (Bezdek, 1980; Hall et al., 1992; Jiang et al., 2014; Qian et al., 2015, 2016, 2018; Zheng et al., 2015; Jiang et al., 2019; Jiang et al., 2020). Cluster analysis is a technique to study the relationship between sample data points. The clustering results reveal the internal connections and differences between sample data and provide an important basis for data processing. The cluster analysis process is shown in **Figure 2**.

The clustering object refers to the grayscale collection of image pixels. The similarity measure is a key step used to measure the similarity between data points. Euclidean distance is often used as the similarity measure. The clustering algorithm uses similarity measures to select data points that are very close to each other and classify them into a cluster. Additionally, the result of clustering will affect the selection of subsequent objects and the similarity measurement, so there is a feedback process. The last step of clustering is to evaluate the clustering results. The importance of evaluating the clustering results will not be lower than the importance of the previous steps. For example, in a cluster analysis method based on a threshold value, it is necessary to judge the classification result to guide the selection of the threshold value. In the cluster analysis method with the function as the criterion, the judgment or evaluation of the classification

TABLE 2 | Related attributes of main brain tissues in MRI images.

Tissue	T1 (s)	T2 (ms)	ρ (1.5)
CSF	0.8–20	110–2,000	70–230
WM	0.76–1.08	61–100	70–90
GM	1.09–2.15	61–109	85–125
Meninges	0.5–2.2	50–165	5–44
Muscle	0.95–1.82	20–67	45–90
Adipose	0.2–0.75	53–94	50–100

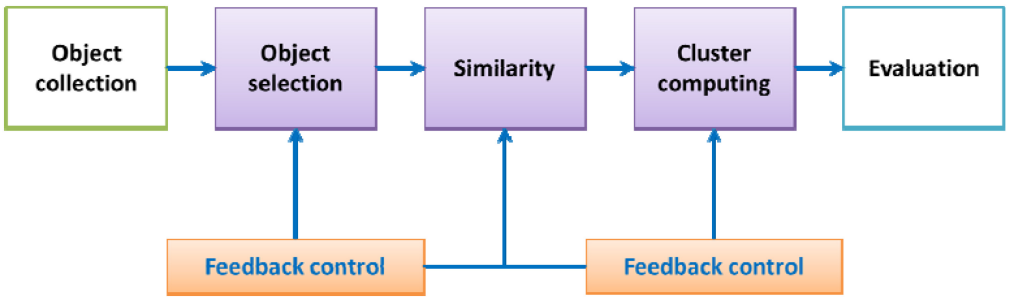


FIGURE 2 | Cluster analysis flowchart.

result can be used to guide the selection of the appropriate criterion function.

Clustering algorithms can be divided into partition-based (Jing et al., 2007; Zhu et al., 2009; Yu et al., 2012; Qian et al., 2017), hierarchy-based (Hall and Goldgof, 2011; Devi and Setty, 2018), density-based (Sunjana and Azizah, 2020; Yin et al., 2020), grid-based (Wang et al., 2019; Zhang et al., 2020), and model-based (Lee et al., 2018; Chrobak et al., 2020) algorithms. Due to the complexity of brain medical images, the boundaries and boundaries of different tissues in the brain are uncertain and unclear. This requires a method for reducing the unclearness. Fuzzy mathematics based on fuzzy sets can solve these uncertain and imprecise problems. The fuzzy clustering algorithm in the clustering algorithm is most suitable for B-MRI-IS. The typical fuzzy clustering algorithm is the FCM algorithm. The principle of image segmentation based on FCM is to use the uncertainty of the classification of image pixels. The degree of membership is used to describe this uncertainty, the distance relationship between pixels is described according to the objective function, and the best cluster center is selected. To describe the clustering algorithm more clearly, **Table 3** shows the symbols used in the FCM algorithm.

Assume the image pixel dataset is $X = \{x_1, x_2, \ldots, x_N\}$, where x_i represents the gray value of the image pixel. FCM's idea is to transform the image segmentation process into the optimization process of the feature function when the objective function converges to realize the fuzzy classification of pixel data. That is, the image segmentation problem is transformed into a clustering problem of dividing N pixels into C classes, and the cluster center of each class is expressed as $Z = \{z_1, z_2, \ldots, z_N\}$. The functional expression of FCM is

$$J_{FCM} = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} d^2\left(x_j, z_i\right) \tag{1}$$

$$d^2\left(x_j, z_i\right) = \left|\left|x_j - z_i\right|\right|^2 \tag{2}$$

The size of m determines the clustering ambiguity. If the value of m is too large, the fuzziness of clustering will increase accordingly, which is not conducive to the defuzzification of the model. In contrast, the smaller the value of m is, the smaller the clustering ambiguity, and the segmentation result is close to hard segmentation. In practical applications, the value of m is usually taken as 2. Equations (3) and (4) give the iterative formula of cluster center z_i and membership degree u_{ij}.

$$z_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_j}{\sum_{j=1}^{N} u_{ij}^{m}} \tag{3}$$

$$u_{ij} = \frac{1}{\sum_{r=1}^{C}\left[\left(\frac{d(x_j, z_i)}{d(x_j, z_r)}\right)^{\frac{2}{m-1}}\right]} \tag{4}$$

The steps of the FCM-based image segmentation algorithm are as follows: (1) Input the brain MRI image; (2) set the fuzzy factor m = 2, the maximum number of iterations ε (ε > 0), the number of clusters C, and initialize the cluster center and membership matrix randomly; (3) according to Eqs 3 and 4, update the clustering center z and membership degree u; (4) make the objective function converge until the cluster center no longer changes; and (5) output the segmented image. The flow of B-MRI-IS based on FCM is shown in **Figure 3**.

TABLE 3 | Description of related symbols in FCM.

Symbols	Description	Symbols	Description
X	Dataset	x_j	The jth pixel
C	Number of clusters	N	Total number of image pixels
u_{ij}	The membership degree of pixel x_j belonging to the jth cluster	z_i	The ith cluster center
d_{ij}	Euclidean distance from sample point x_j to cluster centerz_i	m	Fuzzy factor

Multiview Clustering Algorithm

Traditional fuzzy clustering cannot achieve ideal results when processing multiview data. Based on this requirement, many multiview-related algorithms have been proposed. The existing multiview-related clustering algorithms include coclustering

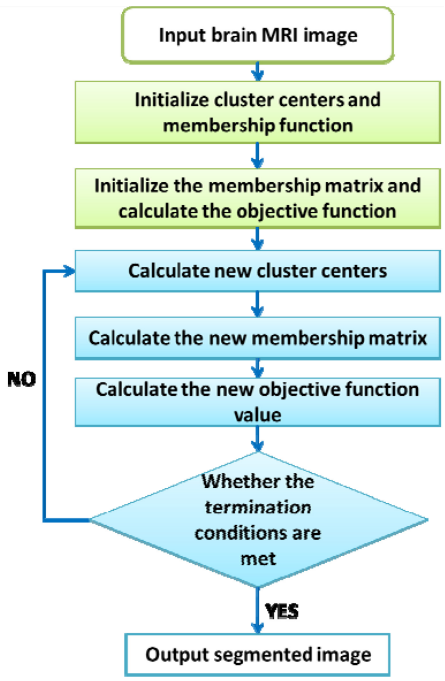


FIGURE 3 | B-MRI-IS process based on FCM.

(Gu and Zhou, 2009), collaborative clustering algorithms (Pedrycz, 2002), and multitask clustering algorithms (Chen et al., 2013). There are many multiview clustering algorithms, the most classic of which is CoFKM. Next, we will briefly introduce the principle and implementation steps of the algorithm. **Table 4** gives the definition of each symbol of the algorithm.

Equation (5) gives the function expression of CoFKM:

$$J_{CoFKM} = \sum_{v=1}^{V}\sum_{i=1}^{C}\sum_{j=1}^{N}\tilde{u}_{ij,v,\eta}\left||x_{j,v} - z_{i,v}\right||^2 \tag{5}$$

$$\tilde{u}_{ij,v,\eta} = (1-\eta)\,u_{ij,v}^{m} + \frac{\eta}{V-1}\sum_{v'=1,v'\neq v}^{V} u_{ij,v'}^{m} \tag{6}$$

$$\begin{cases} u_{ij,v} \in [0,1],\ 1 \le i \le C,\ 1 \le j \le N \\ \sum_{i=1}^{C} u_{ij,v} = 1,\ 1 \le v \le V \end{cases} \tag{7}$$

where $\tilde{u}_{ij,v,\eta}$ represents the weighted fusion of the membership degree $u_{ij,v}^{m}$ under the current view and the membership degree $u_{ij,v'}^{m}$ of the other views. According to Eq. 5, through the Lagrangian extremum solution method, the calculation formulas of the clustering center and membership degree are obtained. Equations (8) and (9) give the calculation formulas of the cluster center and membership matrix of the vth view.

$$z_{i,v} = \frac{\sum_{j=1}^{N}\tilde{u}_{ij,v,\eta}x_{i,v}}{\sum_{j=1}^{N}\tilde{u}_{ij,v,\eta}},\ i = 1, 2, \ldots, C \tag{8}$$

$$u_{ij,v} = \frac{1}{\sum_{h=1}^{C}\left[\frac{(1-\eta)d_{ij,v}^2 + \frac{\eta}{V-1}\sum_{v'=1,v'\neq v}^{V} d_{ij,v'}^2}{(1-\eta)d_{hj,v}^2 + \frac{\eta}{V-1}\sum_{v'=1,v'\neq v}^{V} d_{hj,v'}^2}\right]} \tag{9}$$

Through Eqs 8 and 9, the division matrix for any view is calculated. The final division result is obtained using the ensemble method of the geometric mean. The expression of the integrated method is

$$\widehat{U} = \sqrt[V]{\prod_{v=1}^{V} U_v} \tag{10}$$

Algorithm Evaluation Index

The goal of segmentation is to divide pixels with certain characteristic properties (such as grayscale and spatial position) in an image into a region. All the pixels in a region have similar properties, and the characteristic properties of pixels in different regions are quite different. The evaluation index of the effectiveness of image segmentation is used to describe the pros and cons of the segmentation results. The commonly used evaluation indicators include segmentation accuracy *JS*, Dice coefficient (*DSC*), similarity (*KI*), segmentation coefficient (V_{pc}), segmentation entropy (V_{pe}), and pixel error rate (*ME*). The specific descriptions of commonly used evaluation indicators are shown in **Table 5**.

Experimental Dataset

Common brain image databases are mainly as follows: Brain Perfusion Database, Allen Brain Atlas, BRATS database, simulated brain MR image data from the Brain Web website, and the Internet Brain Segmentation Repository (IBSR). The description of each database is shown in **Table 6**.

TABLE 4 | Description of related symbols in the multiview clustering algorithm.

Symbols	Description	Symbols	Description
$X = \{view_1, view_2, \ldots, view_V\}$	Multiview dataset	$view_v = \{x_{1,v}, x_{2,v}, \ldots, x_{N,v}\}$	vth view data
N	Total number of samples	V	Number of views
$C(2 \le C \le N)$	Number of clusters	η	Parameters that regulate the importance of each view
$Z_v = [z_{1,v}, z_{2,v}, \ldots, z_{c,v}]$	Class center of the vth view	$U_v = [u_{ij,v}]$	Membership of the vth view

TABLE 5 | Evaluation index descriptions.

Index	Description
$JS(S_1, S_2) = \frac{\lvert S_1 \cap S_2 \rvert}{S_1 \cup S_2}$ $DSC(S_1, S_2) = \frac{2\lvert S_1 \cap S_2 \rvert}{\lvert S_1 \rvert + \lvert S_2 \rvert}$	S_1 represents the segmentation area and S_2 represents the ground truth. The larger the *JS* value, the better the segmentation accuracy of the algorithm and the better the effect.
$KI = \frac{TP}{(2TP+FP+FN)/2} = \frac{2TP}{2TP+FP+FN}$	True positive (*TP*), false positive (*FP*), false negative (*FN*); *KI* is a value greater than 0 and less than 1. *KI* = 1 means that the algorithm segmentation result is completely consistent with the standard segmentation. The closer the *KI* is to 1, the more accurate the segmentation result and the better the algorithm performance.
$V_{pc} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{c} u_{ij}^2}{n}$ $V_{pe} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{c} \log u_{ij}}{n}$	These two segmentation indexes are related to the degree of membership u_{ij}. Generally, a segmentation model with less ambiguity is considered to have better performance. The closer V_{pc} is to 1, and the closer V_{pe} is to 0, the higher the segmentation accuracy of the algorithm.
$ME(A, B) = \frac{Area(S_A \cup S_B) - Area(S_A \cap S_B)}{Area(S_B)}$	The segmented image is compared with the "gold standard" image, and the ME. S_A of the obtained model represents the target area extracted by the segmentation algorithm, and S_B represents the "gold standard" target area. The value range of *ME* is 0 to infinity. The closer *ME* is to 0, the smaller the mis-segmentation area and the higher the segmentation accuracy.

TABLE 6 | Brain image database.

Database	Image size	Classification details
Brain Perfusion Database (Cleuziou et al., 2009)	194*237	Number of clusters: 4, which are WM, GM, CSF, and background
Allen Brain Atlas (Okita et al., 2015)	256*128	
BRATS (Rosati et al., 2018)	155*240*240	
Brain Web (Weijer and Gevers, 2006)	181*217*181	
IBSR (Liu et al., 2020)	256*256	

"" Represents multiplication.*

BRAIN MRI IMAGE SEGMENTATION BASED ON MULTIVIEW FUZZY CLUSTERING

Brain MRI Image Segmentation Process Based on Multiview Clustering

MVD carries more information than single-view data. This MVD feature is beneficial to the improvement of cluster analysis. For clustering analysis of MVD, the traditional single-view clustering algorithm is to first separate each view data from the MVD. Second, the single-view clustering algorithm is used to process the data of each view, and the clustering results of the data of each view are obtained. Finally, a suitable ensemble strategy is used to fuse the clustering results of all views to obtain the final division result. The single-view clustering algorithm for processing MVD has the following shortcomings. When the clustering results of a certain view have obvious deviations or the clustering results between different views are very different, if each view is artificially separated for analysis in this way, the final division result obtained by integrated learning is likely to be poor or the performance of the algorithm is unstable. To make full use of the information carried by MVD and improve the accuracy of clustering, this research introduces multiview technology into the single-view clustering method. The multiview clustering algorithm enables the collaborative learning of various views in the clustering process, makes full use of the data information of each view, and improves the clustering performance and stability of the algorithm. The B-MRI-IS process based on multiview clustering is shown in **Figure 4**.

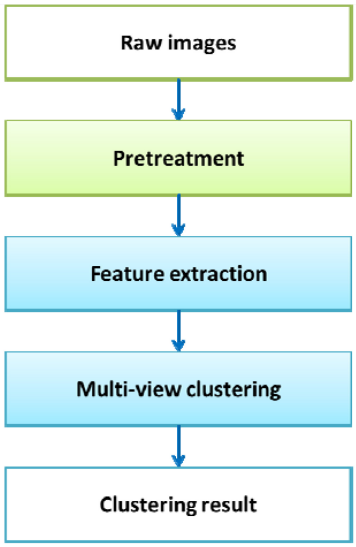


FIGURE 4 | B-MRI-IS process based on multiview clustering.

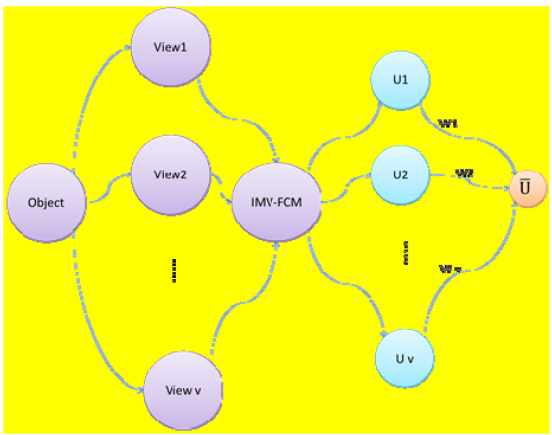


FIGURE 5 | Schematic diagram of IMV-FCM algorithm.

As shown in **Figure 4**, first, the original brain image is preprocessed, such as denoising. Second, the HOG feature, entropy feature, gradient feature, and contrast feature of the original image are extracted. The feature data serve as view data. Third, all view data are input into the multiview clustering model to obtain the final clustering result. Due to the different amounts of information contained in different features, the contribution of the formed view data to the clustering result is also different. Therefore, how to effectively adjust the weight of each view is very important during the use of the multiview clustering algorithm.

IMV-FCM

The CoFKM introduced in the section "Multiview Clustering Algorithm" assigns the same weight to each view, which shows that the CoFKM believes that the contribution of each view data to the clustering is the same. This is obviously inconsistent with the actual situation. For MVD in real production and living environments, the contribution of each view data to the clustering is necessarily different. If the angle of view data with a high degree of contribution is given a high weight and the angle of view data with a low degree of contribution is given a low weight, the overall clustering performance can be improved. Therefore, a multiview fuzzy clustering method with adaptive viewing angle adjustment capability is used in this study, and its objective function is

$$
\begin{aligned}
&J = \sum_{v=1}^{V}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{t=1}^{V} w_{v,t}u_{ij,t}^{m}d_{ij,v}^{2} + \gamma\sum_{v=1}^{V}\sum_{t=1}^{V} w_{v,t}\log(w_{v,t}) \\
&\sum_{i=1}^{C} u_{ij,v} = 1,\ u_{ij,v} \in [0,1] \\
&1 \le i \le C, 1 \le j \le N, 1 \le v \le V \\
&\sum_{t=1}^{V} w_{v,t} = 1,\ w_{v,t} \in [0,1], 1 \le v \le V
\end{aligned}
\tag{11}
$$

where $d_{ij,v}^{2} = \left|\left|x_{j,v} - z_{i,v}\right|\right|^{2}$ and $U_v = \left[u_{ij,v}\right]$ is the division matrix corresponding to the ith view. $\tilde{u}_{ij,v} = \sum_{t=1}^{V} w_{v,t}u_{ij,t}^{m}$ is the division fusion item, which realizes the view fusion of different views in the vth view clustering task. At this time, the importance of each view is reflected by the weight $w_{v,t}$. $w_{v,t}$ represents the importance of the tth view in the vth view cluster. W is the weight matrix of all views. The Lagrange multiplier method is used to solve the extreme value of Eq. 11, and the expression of each

TABLE 7 | Parameter setting of various algorithms based on grid search.

Algorithm	Parameter setting range
FCM	Fuzzy factor m = {1.05,1.1,1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,2}
CoFKM	Fuzzy factor m = {1.05,1.1,1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,2}, collaborative learning parameters $\eta = [0, (K-1)/K]$, K is the number of views
TW-k-means	Regularization parameters $\lambda = \{1, 2, \ldots, 30\}$ and $\eta = \{10, 20, \ldots, 100\}$
CombKM	None
Coclustering	Regularization parameter $\lambda = \{1, 10, 100, 300, 500, 800, 1000\}$, regularization parameter $\mu = \{1, 10, 100, 300, 500, 800, 1000\}$, feature category number $m = \lfloor d/2 \rfloor$, d is the characteristic number, $\lfloor\rfloor$ means rounding down.
IMV-FCM	Fuzzy factor m = {1.05,1.1,1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,2} Regularization parameters $\gamma = \{2^{-12}, 2^{-11}, \ldots, 2^{12}\}$

TABLE 8 | JS indicators of each algorithm (100%).

Noise	Organization	FCM	CoFKM	TW-k-means	CombKM	Coclustering	IMV-FCM
0%	A	85.73%	88.22%	87.76%	87.85%	89.96%	90.45%
	B	74.52%	77.85%	76.90%	77.11%	79.93%	79.76%
	C	65.98%	68.31%	67.58%	68.32%	69.94%	70.74%
	Mean	75.41%	78.13%	77.41%	77.76%	79.94%	80.32%
3%	A	84.98%	87.52%	88.65%	87.03%	87.67%	89.20%
	B	73.87%	76.48%	76.77%	75.96%	76.35%	78.16%
	C	64.76%	66.54%	67.12%	65.87%	66.11%	67.91%
	Mean	74.54%	76.85%	77.51%	76.29%	76.71%	78.42%
5%	A	84.17%	86.11%	86.59%	85.90%	85.97%	87.79%
	B	73.25%	75.43%	75.86%	74.97%	74.98%	76.58%
	C	64.88%	66.35%	66.36%	65.89%	66.99%	67.02%
	Mean	74.10%	75.96%	76.27%	75.59%	75.98%	77.13%
7%	A	82.21%	84.46%	82.32%	82.98%	82.12%	85.80%
	B	72.62%	74.02%	73.52%	73.87%	73.48%	75.35%
	C	64.14%	66.62%	65.07%	65.85%	65.17%	67.56%
	Mean	72.99%	75.03%	73.64%	74.23%	73.59%	76.24%
9%	A	76.68%	82.23%	80.61%	81.02%	81.43%	83.56%
	B	70.25%	73.18%	71.17%	71.65%	71.98%	72.20%
	C	62.27%	65.83%	64.59%	64.04%	64.47%	71.80%
	Mean	69.73%	73.75%	72.12%	72.24%	72.63%	75.85%

TABLE 9 | DSC indicators of each algorithm (100%).

Noise	Organization	FCM	CoFKM	TW-k-means	CombKM	Coclustering	IMV-FCM
0%	A	93.24%	94.75%	94.28%	93.39%	94.02%	95.27%
	B	86.51%	87.79%	87.61%	86.87%	87.99%	88.51%
	C	80.45%	82.44%	82.83%	82.10%	83.04%	85.06%
	Mean	86.73%	88.33%	88.24%	87.45%	88.35%	89.61%
3%	A	92.28%	93.68%	93.77%	93.91%	94.23%	95.08%
	B	85.30%	86.84%	87.04%	86.89%	86.57%	87.62%
	C	79.47%	82.02%	83.22%	82.16%	83.01%	84.41%
	Mean	85.68%	87.51%	88.01%	87.65%	87.94%	89.04%
5%	A	91.36%	91.03%	91.14%	91.09%	91.74%	93.86%
	B	84.89%	84.47%	84.66%	84.82%	84.96%	86.68%
	C	78.55%	80.10%	80.23%	80.34%	80.08%	83.71%
	Mean	84.93%	85.20%	85.34%	85.41%	85.59%	88.08%
7%	A	89.62%	90.02%	90.91%	90.54%	90.74%	92.95%
	B	84.01%	83.36%	83.82%	83.57%	84.10%	85.45%
	C	77.86%	79.03%	79.65%	79.66%	80.12%	82.79%
	Mean	83.83%	84.14%	84.79%	84.59%	84.99%	87.06%
9%	A	85.76%	88.44%	88.62%	88.79%	89.76%	91.87%
	B	83.79%	81.45%	81.89%	81.63%	82.30%	83.67%
	C	77.23%	77.72%	77.79%	77.87%	78.31%	81.32%
	Mean	82.26%	82.54%	82.77%	82.76%	83.46%	85.62%

variable is obtained as follows:

$$z_{i,v} = \frac{\sum_{j=1}^{N}\sum_{t=1}^{V} w_{v,t} u_{ij,t}^{m} x_{j,v}}{\sum_{j=1}^{N}\sum_{t=1}^{V} w_{v,t} u_{ij,t}^{m}} \tag{12}$$

$$u_{ij,t} = \frac{1}{\sum_{h=1}^{C}\left[\frac{\sum_{v=1}^{V} w_{v,t} d_{ij,v}^{2}}{\sum_{v=1}^{V} w_{v,t} d_{hj,v}^{2}}\right]^{1/(m-1)}} \tag{13}$$

$$w_{v,t} = \frac{\exp\left[\frac{-\sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij,t}^{m} d_{ij,v}^{2}}{\gamma}\right]}{\sum_{g=1}^{V}\exp\left[\frac{-\sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij,g}^{m} d_{ij,v}^{2}}{\gamma}\right]} \tag{14}$$

The following ensemble method is used to obtain the final partition matrix:

$$\bar{U} = \max\left(\sum_{t=1}^{V}\left(\frac{\sum_{v=1}^{V} w_{v,t}}{\sum_{t=1}^{V}\sum_{v=1}^{V} w_{v,t}} U_t\right)\right) \tag{15}$$

Figure 5 shows the schematic diagram of the IMV-FCM algorithm. As shown in **Figure 5**, after the multiview data with v views are executed by the IMV-FCM algorithm, v membership matrices are obtained. Each membership matrix is assigned a corresponding weight w, and the final global membership matrix is obtained according to the fusion strategy of Eq. 15.

The specific steps of the IMV-FCM algorithm are in **Algorithm 1**.

Algorithm 1. IMV-FCM algorithm.

Input	Multiview sample set X, number of views V, number of clustering C, iteration threshold ε , fuzzy index m, iteration number I, parameter γ .
Output	The final division matrix $\bar{U}$, the clustering center of each view $z_{i,v}$, the view fusion weight matrix $W = \{w_{v,t}\}$.
Step 1	Randomly generate fuzzy membership matrix $u_{ij,t}$ $(1 \le t \le V)$ for each view and view fusion weight matrix $W = \{w_{v,t}\}$ for each view;
Step 2	According to Eq. 12, the cluster center $z_{i,v}$ of each view is updated.
Step 3	According to Eq. 13, the membership degree $u_{ij,t}$ of each view is updated.
Step 4	According to Eq. 14, the view fusion weight matrix $W = \{w_{v,t}\}$ is updated.
Step 5	If $\|\|J^{l+1} - J^{l}\|\| < \varepsilon$, the algorithm stops iterating; otherwise, it returns to Step 2.
Step 6	After the algorithm converges, the fuzzy membership of each view is output.
Step 7	According to the fuzzy membership degree of each view obtained in Step 6, Eq. 15 is used to obtain the final division matrix.

SIMULATION EXPERIMENT ANALYSIS

Experimental Background

The dataset used in this study was the data downloaded from BrainWeb. The 2D images of 89 cross sections (89 CS), 92 cross sections (95 CS), and 95 cross sections (95 CS) in the T1-weighted brain MRI image were selected for segmentation. The comparison algorithms used were FCM, CoFKM (Cai et al., 2019), two-layer automatic weighted

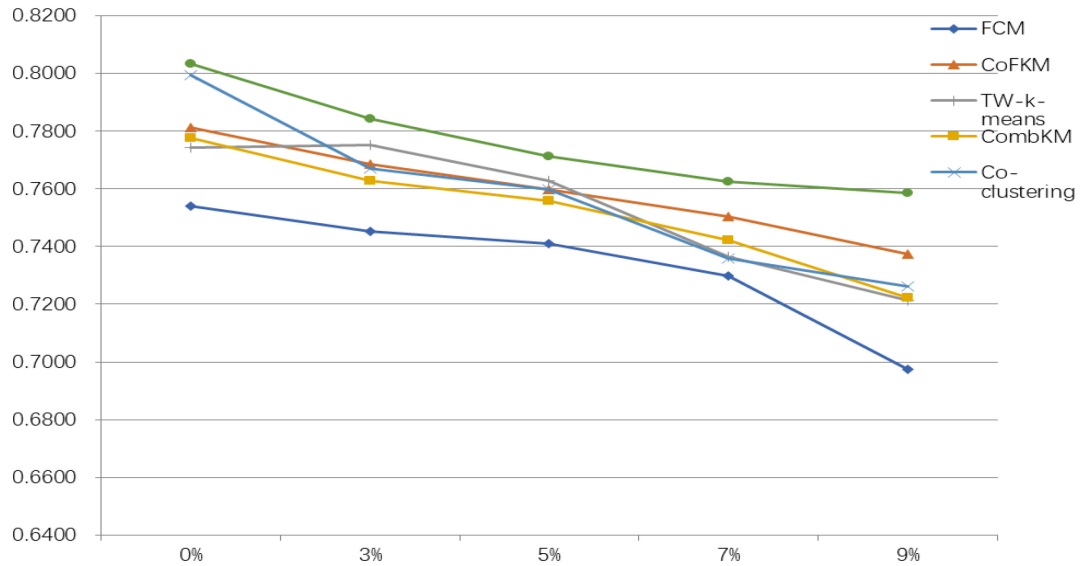


FIGURE 6 | JS indicator drop rate of each algorithm.

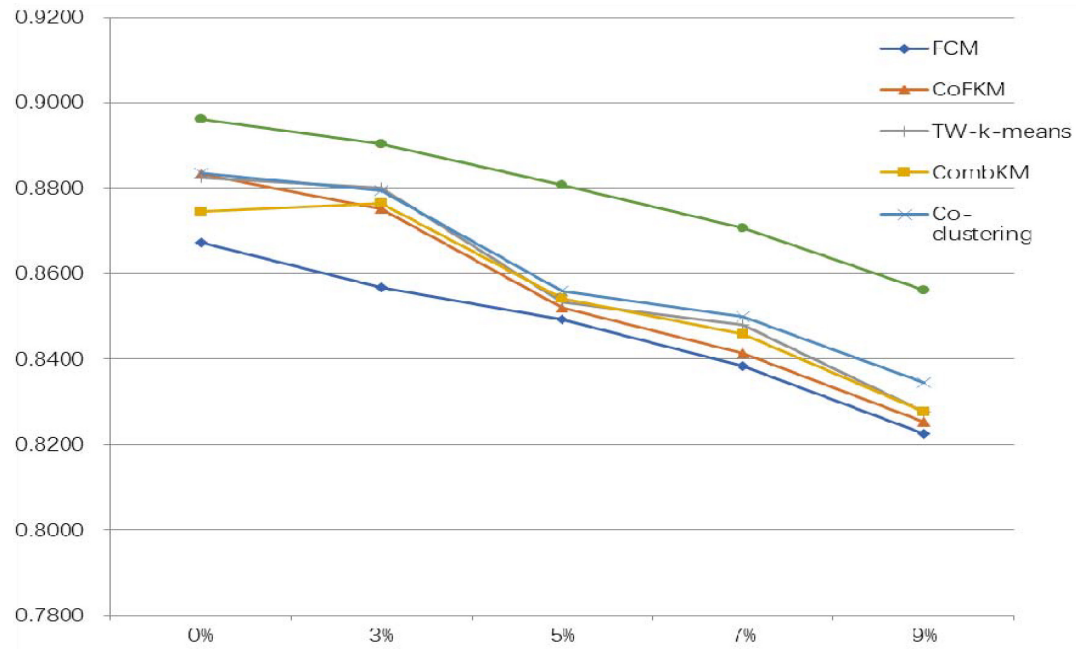


FIGURE 7 | DSC index drop rate of each algorithm.

clustering algorithm (TW-k-means) (Singh et al., 2020), multitask-based K-means (CombKM) (Chen et al., 2013), and collaborative clustering based on sample and feature space (coclustering) (Gu and Zhou, 2009). In the experiment, the iteration stop threshold ε of each algorithm is set to 0.001, and the maximum number of iterations l was set to 100. The parameter settings of each algorithm are shown in **Table 7**. The parameter selection methods of each comparison algorithm include the grid search method and clustering index. This article mainly uses the grid search method to confirm the optimal parameters. The reason for choosing the grid search method for parameter selection is that the performance of the model trained based on the parameters selected by this method is the best. The evaluation index adopts JS and DSC. All experiments were based on the MATLAB 2018a programming environment, which was simulated on a Lenovo PC configured with a Windows 10 operating system, 2.60 GHz CPU, and 8G memory Intel(R) Core(TM)i7 processor.

Experimental Results and Analysis

To verify the superiority of the IMV-FCM algorithm, the experiment section gives a comparison of the effect of each algorithm on image segmentation. In analyzing the antinoise performance of the algorithm, this study adds different proportions of Gaussian noise to the image. The segmentation results of each algorithm are shown in **Tables 8**, **9**. The symbol A represents WM, B represents GM, and C represents CSF.

From the experimental data in **Tables 8**, **9**, it can be determined that when processing original brain images, the segmentation performance of the multiview algorithm used in this study is significantly better than FCM. This shows that the introduction of the multiview mechanism can mine information from multiple views, thereby improving the clustering accuracy. Compared with other multiview clustering algorithms, the segmentation effect of the IMV-FCM algorithm in this study is better. This shows that the IMV-FCM algorithm with partition adaptive fusion capability requires the Co-FKM algorithm and CombKM algorithm to manually set the degree of partition fusion. The IMV-FCM algorithm can obtain better view weights during multiview learning and finally obtain better multiview clustering effects.

As the image noise increases, the clustering performance of all clustering algorithms begins to decline. However, even if the performance of each algorithm is declining, the performance of the IMV-FCM algorithm is still better than that of the other algorithms. This is manifested in two aspects. First, the clustering performance of IMV-FCM is better than that of the comparison algorithm. The second is that with the increase in image noise, the clustering performance of the IMV-FCM algorithm decreases the most slowly, as shown in **Figures 6**, **7**. This fully shows that IMV-FCM is more robust to noise.

CONCLUSION

To further improve the performance of B-MRI-IS, the IMV-FCM algorithm is applied in this research. The IMV-FCM algorithm is a multiview fuzzy clustering algorithm with adaptive view weight. Under the adaptive learning of the view fusion weight matrix, the coordination between various views is more flexible. Additionally, each view can be self-adapted to learn and then achieve a better clustering effect. Since the original brain MRI image is not MVD, this study extracts multiple feature data of the original image and uses one feature data as one view data to construct MVD. The experimental results prove that the segmentation method used in this paper optimizes the segmentation effect. With the increase in noise, the clustering effect of IMV-FCM decreases more slowly than the comparison method, which shows that the method used has stronger noise immunity. Although the effectiveness of IMV-FCM was verified, it still faces certain limitations. For example, the proposed method uses the classic FCM algorithm as a framework and uses Euclidean distance, which causes it to suffer from the dimensionality disaster problem when facing high-dimensional multiview clustering tasks. How to solve this problem will be the focus of future research. In addition, similar to most unsupervised learning algorithms, the selection of optimal parameters is an important issue. Different practical applications require different ranges of parameter values. Since the optimal parameters are usually determined by the application, for an exact application, the appropriate parameter range of the algorithm used can be determined through prior knowledge or available valid labeled datasets. The ensemble learning strategy can avoid the selection of optimal parameters to a certain extent, and it is planned to be further studied in follow-up work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://brainweb.bic.mni.mcgill.ca.

AUTHOR CONTRIBUTIONS

LH and YG conceived and developed the theoretical framework of the manuscript. YG and XG designed and performed the surgical procedure of the case presentation. LH, JX, and TN performed the data evaluation, analysis, and interpretation, designed the figures, designed and wrote the manuscript in consultation with YG and XG, who took the lead in writing the manuscript. All authors participated in the editing process.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grants 81701793, 61702225, 61772241, 61873321, and61806026; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180956; in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127; in part by the Science and Technology Demonstration Project of Social Development of Wuxi under Grant WX18IVJN002; and in part by the Youth Foundation of the Commission of Health and Family Planning of Wuxi under Grant Q201654.

REFERENCES

Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., and Moriarty, T. (2002). Amodified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Trans. Med. Imaging* 21, 193–199. doi: 10.1109/42.996338

Bezdek, J. C. (1980). A convergence theorem for the fuzzy ISODATA clustering algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* 2, 1–8. doi: 10.1109/TPAMI.1980.4766964

Bezdek, J. C. (1981). *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY: Plenum Press.

Cai, T. T., Ma, J., and Zhang, L. (2019). Chime: clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality. *Ann. Stat.* 47, 1234–1267. doi: 10.1214/18-AOS1711

Chen, S., and Zhang, D. (2004). Robust image segmentation using FCM with spatialconstraints based on new kernel-induced distance measure. *IEEE Trans. Syst. Man Cybern. B Cyber.* 34, 1907–1916. doi: 10.1109/TSMCB.2004.831165

Chen, X., Xu, X., Huang, J. Z., and Ye, Y. (2013). TW-k-means: automated two-level variable weighting clustering algorithm for multiview data. *IEEE Trans. Knowl. Data Eng.* 25, 932–944. doi: 10.1109/TKDE.2011.262

Chrobak, M., Dürr, C., Aleksander, F., and Nilsson, B. J. (2020). Online clique clustering. *Algorithmica* 82, 938–965. doi: 10.1007/s00453-019-00625-1

Chuang, K. S., Tzeng, H. L., Chen, S., Wu, J., and Chen, T. J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *Comput. Med. Imaging Graph. Official J. Comput. Med. Imaging Soc.* 30, 9–15. doi: 10.1016/j.compmedimag.2005.10.001

Cleuziou, G., Exbrayat, M., Martin, L., and Sublemontier, J. H. (2009). "Co FKM: a centralized method for multiple-view clustering," in *Proceedings of the 9th IEEE Internaional Conference on Data Mining*, (Miami FL: IEEE), 752–757. doi: 10.1109/ICDM.2009.138

Devi, B. R., and Setty, S. P. (2018). Hybrid clustering algorithm 'KCu' for combining the features of K-means and CURE Algorithm for efficient outliers handling. *Adv. Model. Anal. B* 61, 76–79. doi: 10.18280/ama_b.610204

Elazab, A., Wang, C., Jia, F., Wu, J., Li, G., and Hu, Q. (2015). Segmentation of brain tissues from magnetic resonance images using adaptively regularized kernel-based fuzzy c-means clustering. *Comput. Math. Methods Med.* 2015:485495. doi: 10.1155/2015/485495

Gong, M., Liang, Y., Shi, J., Ma, W., and Ma, J. (2013). Fuzzy c-means clustering with local information and kernel metric for image segmentation. *IEEE Trans. Image Process.* 22, 573–584. doi: 10.1109/TIP.2012.2219547

Gu, J., Cheng, T., Hua, L., Wang, J., Zhao, J., and Cao, Y. (2019). Overview of image segmentation and registration for spine biological modeling. *J. Syst. Simul.* 31, 167–173. doi: 10.16182/j.issn1004731x.joss.18-0806

Gu, Q., and Zhou, J. (2009). Learning the shared subspace for multi-task clustering and transductive transfer classification. *Proceedings of the 2009 9th IEEE International Conference on Data Mining*, (Miami Beach, FL: IEEE), 159–168. doi: 10.1109/ICDM.2009.32

Hall, L. O., Bensaid, A. M., and Clarke, L. P. (1992). A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE Trans. Neural Netw.* 3, 672–681. doi: 10.1109/72.159057

Hall, L. O., and Goldgof, D. B. (2011). Convergence of the single-pass and online fuzzy C-means algorithms. *IEEE Trans Fuzzy Syst.* 19, 792–794. doi: 10.1109/TFUZZ.2011.2143418

Ji, Z. X., Sun, Q. S., and Xia, D. S. (2011). A modified possibilisticfuzzy c-means clustering algorithm for bias field estimation and segmentation of brain MR image. *Comput. Med. Imaging Graph.* 35, 383–397. doi: 10.1016/j.compmedimag.2010.12.001

Jiang, C. F., Chang, C. C., and Huang, S. H. (2012). Regions of interest extraction fromspect images for neural degeneration assessment using multimodality image fusion. *Multidimen. Syst. Signal Process.* 23, 437–449. doi: 10.1007/s11045-011-0162-3

Jiang, Y., Bi, A., Xia, K., Xue, J., and Qian, P. (2020). Exemplar-based data stream clustering toward internet of things. *J. Supercomput.* 76, 2929–2957. doi: 10.1007/s11227-019-03080-5

Jiang, Y., Chung, F. L., Wang, S., Deng, Z., Wang, J., and Qian, P. (2014). Collaborative fuzzy clustering from multiple weighted views. *IEEE Trans. Cybern.* 45, 688–701. doi: 10.1109/TCYB.2014.2334595

Jiang, Y., Zhao, K., Xia, K., Xue, J., Zhou, L., Ding, Y., et al. (2019). A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation. *J. Med. Syst.* 43:118. doi: 10.1007/s10916-019-1245-1

Jing, L. P., Ng, M. K., and Huang, J. Z. (2007). An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data. *IEEE Trans. Knowl. Data Eng.* 19, 1026–1041. doi: 10.1109/TKDE.2007.1048

Krinidis, S., and Chatzis, V. (2010). A robust fuzzy local information C-means clustering algorithm. *IEEE Trans. Image Process.* 19, 1328–1337. doi: 10.1109/TIP.2010.2040763

Kumar, S. N., Lenin, F. A., and Sebastin, V. P. (2020). An overview of segmentation algorithms for the analysis of anomalies on medical images. *J. Intell. Syst.* 29, 612–625. doi: 10.1515/jisys-2017-0629

Lee, K., Moon, C., and Nam, Y. (2018). Diagnosing vocal disorders using cobweb clustering of the jitter, shimmer, and harmonics-to-noise ratio. *KSII Trans. Internet Inform. Syst.* 12, 5541–5554. doi: 10.3837/tiis.2018.11.020

Li, C., Gore, J. C., and Davatzikos, C. (2014). Multiplicative intrinsic component optimization(MICO) for MRI bias field estimation and tissue segmentation. *Magn. Reson. Imaging* 32, 413–439. doi: 10.1016/j.mri.2014.03.010

Liu, L., Kuang, L., and Ji, Y. (2020). Multimodal MRI brain tumor image segmentation using sparse subspace clustering algorithm. *Comput. Math. Methods Med.* 2020:8620403. doi: 10.1155/2020/8620403

Okita, Y., Miyata, A. H., Motomura, B. N., and Takamoto, S. (2015). A study of brain protection during total arch replacement comparing antegrade cerebral perfusion versus hypothermic circulatory arrest, with or without retrograde cerebral perfusion: analysis based on the Japan adult cardiovascular surgery database. *J. Thorac. Cardiovasc. Surgery* 149, 65–73. doi: 10.1016/j.jtcvs.2014.08.070

Pedrycz, W. (2002). Collaborative fuzzy clustering. *Pattern Recogn. Lett.* 23, 1675–1686.

Qian, P., Jiang, Y., Deng, Z., Hu, L., Sun, S., Wang, S., et al. (2015). Cluster prototypes and fuzzy memberships jointly leveraged cross-domain maximum entropy clustering. *IEEE Trans. Cybern.* 46, 181–193. doi: 10.1109/TCYB.2015.2399351

Qian, P., Jiang, Y., Wang, S., Su, K. H., Wang, J., Hu, L., et al. (2016). Affinity and penalty jointly constrained spectral clustering with all-compatibility, flexibility, and robustness. *IEEE Trans. Neural Netw. Learn. Syst.* 28, 1123–1138. doi: 10.1109/TNNLS.2015.2511179

Qian, P., Zhao, K., Jiang, Y., Su, K. H., Deng, Z., Wang, S., et al. (2017). Knowledge-leveraged transfer fuzzy C-Means for texture image segmentation with self-adaptive cluster prototype matching. *Knowl. Based Syst.* 130, 33–50. doi: 10.1016/j.knosys.2017.05.018

Qian, P., Zhou, J., Jiang, Y., Liang, F., and Muzic, R. F. Jr. (2018). Multi-view maximum entropy clustering by jointly leveraging inter-view collaborations and intra-view-weighted attributes. *IEEE Access* 6, 28594–28610. doi: 10.1109/ACCESS.2018.2825352

Rosati, P., Lupascu, C. A., and Tegolo, D. (2018). Analysis of low-correlated spatial gene expression patterns: a clustering approach in the mouse brain data hosted in the Allen Brain Atlas. *IET Comput. Vision* 12, 996–1006. doi: 10.1049/iet-cvi.2018.5217

Singh, M., Vishal, V., Verma, A., and Sharma, N. (2020). Segmentation of MRI data using multi-objective antlion based improved fuzzy c-means. *Biocybern. Biomed. Eng.* 40, 1250–1266. doi: 10.1016/j.bbe.2020.07.001

Sunjana and Azizah, Z. (2020). Outlier detection of transaction data using DBSCAN algorithm. *Int. J. Psychosoc. Rehabil.* 24, 3232–3240. doi: 10.37200/IJPR/V24I2/PR200632

Thanh, M. N., and Wu, Q. M. J. (2013). A fuzzy logic model based Markov random field for medical image segmentation. *Evolv. Syst.* 4, 171–181. doi: 10.1007/s12530-012-9066-1

Wang, J., Schreiber, D. K., Bailey, N., Hosemann, P., and Toloczko, M. B. (2019). The application of the OPTICS algorithm to cluster analysis in atom probe tomography data. *Microsc. Microanal.* 25, 338–348. doi: 10.1017/S1431927618015386

Wang, X. Y., Zhang, D. D., and Wei, N. (2015). Fractal image coding algorithm using particle swarm optimisation and hybrid quadtree partition scheme. *Iet Image Process.* 9, 153–161. doi: 10.1049/iet-ipr.2014.0001

Weijer, J., and Gevers, T. (2006). Boosting color saliency in image feature. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 150–156. doi: 10.1109/TPAMI.2006.3

Yin, S., Li, H., Liu, D., and Karim, S. (2020). Active contour modal based on density-oriented BIRCH clustering method for medical image segmentation. *Multimedia Tools Appl.* 79, 31049–31068. doi: 10.1016/j.ijleo.2018.01.004

Yu, S., Tranchevent, L., Liu, X., Glanzel, W., Suykens, J. A. K., De Moor, B., et al. (2012). Optimized data fusion for kernel k-means clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 1031–1039. doi: 10.1109/TPAMI.2011.255

Zhang, C., Churazov, E., and Zhuravleva, I. (2020). Pairs of giant shock waves (N-waves) in merging galaxy clusters. *Monthly Notices R. Astronomic. Soc.* 501, 1038–1045.

Zheng, Y., Byeungwoo, J., Xu, D., Wu, Q. M. J., and Hui, Z. (2015). Image segmentation by generalized hierarchical fuzzy C- means algorithm. *J. Intell. Fuzzy Syst.* 28, 4024–4028. doi: 10.3233/IFS-141378

Zhu, L., Chung, F. L., and Wang, S. T. (2009). Generalized fuzzy k-means clustering algorithm with improved fuzzy partitions. *IEEE Trans Syst. Man Cybern.* 39, 578–591. doi: 10.1109/TSMCB.2008.2004818

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 April 2021
doi: 10.3389/fnins.2021.651762

Spiking Neural Network for Augmenting Electroencephalographic Data for Brain Computer Interfaces

Sai Kalyan Ranga Singanamalla[1] *and Chin-Teng Lin*[1,2]*

[1] Computational Intelligence and Brain Computer Interface Lab, School of Computer Science, University of Technology Sydney, Sydney, NSW, Australia, [2] Centre for Artificial Intelligence, University of Technology Sydney, Sydney, NSW, Australia

With the advent of advanced machine learning methods, the performance of brain–computer interfaces (BCIs) has improved unprecedentedly. However, electroencephalography (EEG), a commonly used brain imaging method for BCI, is characterized by a tedious experimental setup, frequent data loss due to artifacts, and is time consuming for bulk trial recordings to take advantage of the capabilities of deep learning classifiers. Some studies have tried to address this issue by generating artificial EEG signals. However, a few of these methods are limited in retaining the prominent features or biomarker of the signal. And, other deep learning-based generative methods require a huge number of samples for training, and a majority of these models can handle data augmentation of one category or class of data at any training session. Therefore, there exists a necessity for a generative model that can generate synthetic EEG samples with as few available trials as possible and generate multi-class while retaining the biomarker of the signal. Since EEG signal represents an accumulation of action potentials from neuronal populations beneath the scalp surface and as spiking neural network (SNN), a biologically closer artificial neural network, communicates via spiking behavior, we propose an SNN-based approach using surrogate-gradient descent learning to reconstruct and generate multi-class artificial EEG signals from just a few original samples. The network was employed for augmenting motor imagery (MI) and steady-state visually evoked potential (SSVEP) data. These artificial data are further validated through classification and correlation metrics to assess its resemblance with original data and in-turn enhanced the MI classification performance.

Keywords: spiking neural network, electroencephalography, brain computer interface, motor imagery, data augmentation

OPEN ACCESS

Edited by:
Mohammad Khosravi, Persian Gulf University, Iran

Reviewed by:
Jamshid Pirgazi, University of Science and Technology of Mazandaran, Iran
Yinglei Song, Jiangsu University of Science and Technology, China

***Correspondence:**
Chin-Teng Lin
chin-teng.lin@uts.edu.au

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *10 January 2021*
Accepted: *22 February 2021*
Published: *01 April 2021*

Citation:
Singanamalla SKR and Lin C-T (2021) Spiking Neural Network for Augmenting Electroencephalographic Data for Brain Computer Interfaces. Front. Neurosci. 15:651762. doi: 10.3389/fnins.2021.651762

1. INTRODUCTION

Brain–computer interfaces (BCIs) are a form of human–computer interaction through which users can communicate with an external device or application, such as wheelchair navigation, playing games, operating prosthetics, or using a keyboard speller, via their thoughts (Donoghue, 2002; Moore, 2003; Schalk et al., 2004). The fundamental principle of a BCI system is interpreting brain signals to extract reliable markers for decoding the user intentions and translate this information

as a command to an external application. In BCI applications, electroencephalography (EEG) is the most commonly used brain imaging method for monitoring brain signals due to its portability and high temporal resolution (Cincotti et al., 2008; Alwasiti et al., 2010; Fazli et al., 2012).

EEG–BCI systems can employ different modalities, such as motor imagery (MI) (Marchesotti et al., 2016), steady-state visually evoked potential (SSVEP) (Nakanishi et al., 2014), and P300 potentials (Jin et al., 2015). The recent developments in deep learning have resulted in an unprecedented improvement in the performance of BCI systems (Zhang et al., 2017; Chiarelli et al., 2018; Schwemmer et al., 2018). However, such machine learning techniques require abundant user-specific labeled data samples for effective training, and recording a large sample size is experimentally a tedious process. Additionally, EEG data are often contaminated with noise and artifacts, leading to the exclusion of many samples during the data pre-processing steps. Generating synthetic EEG samples could aid in the training of machine learning algorithms. Lotte (2011) generated synthetic EEG trials by chunking each original EEG trial into a different segment and then combine the segments from different trials to form a new artificial sample. But this process results in a trade-off with the loss of spectral-domain features in the artificial signal and requires a careful division of segments to maintain the temporal features. In another study, Dinarès-Ferran et al. (2018) employed empirical mode decomposition (EMD), in which a signal is decomposed into a finite number of intrinsic function, a nonlinear oscillatory signal. These decomposed signals from different EEG trials are again composed to produce a new EEG sample. Though this method aims to retain the spectral domain features, the study stated this method could often produce irrelevant artificial samples.

In the recent decade, various deep learning-based generative models were formulated especially for image and speech synthesis. The most prominent of these models include generative adversarial networks (GAN) and its variants, variational autoencoders (VAE), etc. Such models have been explored for generating artificial EEG data with a focus on performance enhancement in BCI applications. For example, Aznan et al. (2019) generated artificial SSVEP data using deep convolution GAN (DCGAN), Wassertian GAN (W-GAN), and VAE. Similarly, Hartmann et al. (2018) used modified W-GANs to stabilize its training process in generated left-hand motor imagery data. However, these generative models can be trained to produce only one class of data. Although GAN-based networks have shown promising output, training the model is complex due to GAN instability, fine-tuning of hyperparameters, and a sufficiently large number of samples are needed for training. And, we believe requiring huge data for generating even more samples is counterintuitive. Therefore, a new generative model is required that can produce multi-class artificial EEG data with as few available original data samples as possible.

Humans process information through ensembles of spiking neurons and the reorganization of collective chaotic neuronal activity to produce varied behaviors and actions remain a key neuroscience research topic (Sussillo and Abbott, 2009; Churchland et al., 2012; Gilra and Gerstner, 2017; Nicola and Clopath, 2017). Nicola and Clopath (2017) enforced such behavior in a recurrent spiking neural network (SNN) using FORCE, a supervised learning method, and reproduced an array of signals such as sinusoidal waves, Lorenz attractor, Ode to Joy, bird song, and movie replay mimicking the hippocampal region. In another study, Ingrosso and Abbott (2019) focused on dynamically balanced recurrent spiking networks for reconstructing motion tracking signals using bounded constrained coordinate descent optimization.

Inspired by these studies, understanding the underlying spiking dynamics in EEG generation could potentially advance BCI systems. Previously adopted methods for signal reconstruction, such as recursive least square (RLS), have been shown to train SNNs with remarkable performance. However, these methods are usually suitable for periodic signals and require finely tuned hyperparameters for training different signals, i.e., the same set of hyperparameters cannot account for training different signals. In contrast, EEG is non-stationary and has high variance. To address these issues, we implemented a feedforward SNN and trained the network via surrogate-gradient descent (Zenke and Ganguli, 2018; Neftci et al., 2019) method to reconstruct EEG template and, in turn, generate synthetic EEG signals.

The non-stationarity of EEG signals and their uniqueness across recorded trials for a given stimulus pose a challenge in reconstructing the signal and in generating a synthetic dataset in which each sample is different. The famous rodent spike train repeatability experiment (Mainen and Sejnowski, 1995) has shown that the same stimulus can elicit different spike trains (with slight changes in spike timings) at neocortical neurons owing to the variability either with unknown information flow from other circuits or intrinsic background noise in the system. Inspired by this concept, we aimed to influence the trained SNN model through an external neural perturbation layer as background noise. By modulating the perturbation layer's noise level, the proposed model was able to (theoretically) generate an unlimited number of samples of MI and SSVEP. One of the main use of synthetic data in machine learning is to improve classifier performance. The synthetic MI-EEG data enhanced the classification accuracy when used for training the classifier. The major contributions of this work are threefold: (1) the development of an SNN that is agnostic to the EEG signal modality, i.e., the same model architecture to accommodate varied EEG signals (e.g., MI and SSVEP) without extensively change the hyperparameters, (2) the generation of multi-class synthetic motor imagery EEG data from only a few original samples, and (3) validating the artificial MI data through classification and performance enhancement.

2. METHODS

A feedforward spiking neural network, as depicted in **Figure 1**, with each node behaving as a Leaky-Integrate and Fire (LIF) neuron model, was designed for producing EEG signals. As the ground truth spike train triggering a targeted EEG signal is inaccessible, a Poisson-generated spike train with a firing rate of

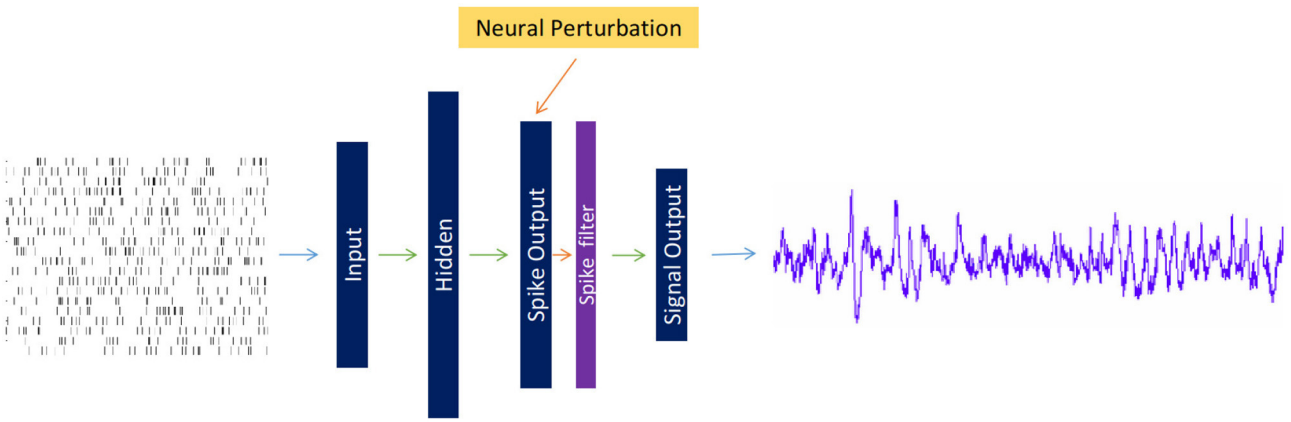


FIGURE 1 | Depiction of a three-layered feedforward spiking network mapping a random spike train to an electroencephalography (EEG) signal with additional neural perturbation layer of Poisson neurons distorting the information of Spike Output layer. The spike train from the Spike Output layer is converted to a smoothed signal via a double exponential spike filter, which is then transformed to an EEG signal. The number of nodes in each layer varies depending on the task for best fit.

10 Hz acts as the input layer preceding with one hidden and one output layer. The number of neurons in each layer is different for MI and SSVEP reconstruction (see **Table 1**). The spiking activity from the output layer is transformed to a rate signal using a double exponential synaptic filter (see Equation 9), which is then weighted averaged to produce an EEG signal. All the weight matrices between layers except for perturbation layers were trained during the reconstruction process. The number of nodes in the signal output layer varies depending on the number of EEG channels that were target toward generation. The perturbation layer containing Poisson neurons acts as background noise whose parameters i.e., the number of noise neurons and its firing rate, are varied to assess the effect of SNN on artificial data.

The entire data generation process (see **Figure 2**) is performed in three stages: (a) extracting reliable EEG templates (one for each class) either through averaging original samples (in SSVEP) or selecting the best trial among all the samples (in MI), (b) training the SNN to reconstruct EEG templates of all classes simultaneously, and (c) activating the perturbation layer after completion of the training process to generate synthetic EEG samples. These stages were performed in the mentioned sequential order. In any recording, EEG signals for a given stimulus vary due to noise, artifacts, and background activity. However, under ideal conditions, all the trials retain the stimulus-related biomarker. Therefore, for stage 1 (i.e., template extraction), we averaged few trials to reduce noise and obtain a clean signal with a good biomarker (a strong power peak at target frequency in the power spectrum) in the case of SSVEP. Contrarily for MI, averaging impaired its biomarker alongside noise as the MI's biomarker spread in a range of mu and beta spectral bands. Therefore, we chose the best trial in a dataset as a template.

To extract a reliable EEG template in MI, a classification algorithm was applied on the original dataset for both training and testing. During testing, the probability of each sample to identify its class or label is estimated. The sample with highest probability (one for each class) was considered as template. For example, if two samples have probability of 0.9 and 0.8 to be labeled into a particular class, then the sample with highest probability (i.e., sample with 0.9 probability) is assumed as reliable template due to its high value.

Using the surrogate-gradient descent method for weight updating, the SNN is initially trained to reconstruct an EEG template. Then, the neural perturbation layer composed of Poisson neurons (acting as background activity) injects random current (a random spike train for each new sample) to the output layer via fixed synaptic weights, enabling the network to generate synthetic EEG samples that differ from the original signal while retaining its biomarker.

TABLE 1 | List of parameters in the model equations, the spiking neural network (SNN) architecture, and the optimizer.

Equation parameters	**value**
dt	1 mS
u_{rest}	0 mV
τ_m	10 mS
τ_s	5 mS
R	1 M
ϑ	1 mV
τ_d	2 mS
τ_r	3 mS
Model parameters	**Spiking layer nodes**
MI	[500, 1500, 100]
SSVEP	[100, 500, 200]
Optimizer properties	**Value**
Learning rate	0.0005
Betas	(0.9, 0.999)
Epsilon	8×10^{-1}
Weight decay	0

Spiking layer nodes represents an array of the number of neurons in each of the spiking network's layer.

2.1. Neuron Model

Spiking neural networks are constructed with neurons that show biological realism. A wide range of neuron models exists with varying complexity, such as the Hodgkin Huxley (HH), the LIF neuron, the FitzHugh-Nagumo model, and the Izhikevich model. Opting for computational simplicity, the LIF neuron model, whose membrane potential is described by Equation (1), was chosen as a basis for constructing each node in a spiking layer.

$$\tau_{\mathrm{m}} \frac{\mathrm{d}u_i}{\mathrm{d}t} = -\left(u_i - u_{\mathrm{rest}}\right) + RI_i \tag{1}$$

$$\frac{\mathrm{d}I_i}{\mathrm{d}t} = -\frac{I_i(t)}{\tau_{\mathrm{s}}} + \sum_j W_{ij} S_j(t) \tag{2}$$

where $u_i(t)$ is the membrane potential of a neuron, i, u_{rest} is the resting potential, τ_m is the membrane time constant, R is the input resistance, $I_i(t)$ is the input current, and τ_s is the synaptic time constant. The membrane potential increases with input current, and after reaching a threshold ϑ, the neuron emits a spike and resets its potential to u_{rest}. Incorporating the reset property into Equation (1):

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = -\frac{1}{\tau_{\mathrm{m}}}\left(u_i - u_{\mathrm{rest}}\right) + RI_i + S_i(t)\left(u_{\mathrm{rest}} - \vartheta\right) \tag{3}$$

$$S_i(t) = \sum_k \delta\left(t - t_i^k\right) \tag{4}$$

$S_i(t)$ is the spike train (sum of a dirac delta function, δ) of a neuron and t_i^k is the kth firing time of the corresponding neuron (see Equation 4). As suggested in Neftci et al. (2019), an approximated version of the LIF membrane potential for a small simulation time step $\Delta_t > 0$, as shown in Equation (6), with the substitution of numerical value of R, u_{rest}, and ϑ from **Table 1**, was implemented in this study.

$$I_i[n+1] = \alpha I_i[n] + \sum_j W_{ij} S_j[n] \tag{5}$$

$$u_i[n+1] = \beta u_i[n] + I_i[n] - S_i[n] \tag{6}$$

where $\alpha \equiv \exp\left(-\frac{\Delta_t}{\tau_{\mathrm{s}}}\right)$, $\beta \equiv \exp\left(-\frac{\Delta_t}{\tau_{\mathrm{m}}}\right)$.

Two other neuron models, the quadratic integrate-and-fire (QIF) model (Equation 7) and the exponential integrate-and-fire (EIF) model (Equation 8), were also tested for EEG reconstruction (see **Supplementary Figures 2, 3**).

$$\tau_{\mathrm{m}} \frac{\mathrm{d}}{\mathrm{d}t} u_i = a_0 \left(u_i - u_{\mathrm{rest}}\right)\left(u_i - u_c\right) + RI_i \tag{7}$$

$$\tau_{\mathrm{m}} \frac{\mathrm{d}}{\mathrm{d}t} u_i = -\left(u_i - u_{\mathrm{rest}}\right) + \Delta_T \exp\left(\frac{u_i - \vartheta}{\Delta_T}\right) + RI_i \tag{8}$$

where, $a_0 > 0$, $u_c > u_{\mathrm{rest}}$ and sharpness parameter $\Delta_T > 0$. And, I_i follows the same dynamics as Equation (2).

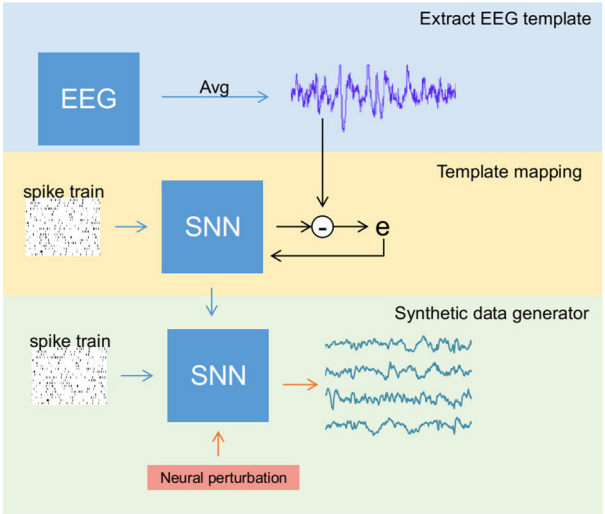


FIGURE 2 | Overview of the steps involved in generating synthetic electroencephalography (EEG) samples using the spiking neural network (SNN). The process is divided into three different stages: (1) Template extraction, where EEG template from the sample data is derived (blue box), (2) training SNN to produce EEG signal template (yellow box), and (3) activating the perturbation layers that help SNN to generate artificial samples (green box).

2.2. Synaptic Filter

Spike trains of the output layer are filtered by a double exponential synaptic filter r_j for each neuron j as shown in the following equation:

$$\begin{aligned} \dot{r}_j &= -\frac{r_j}{\tau_d} + h_j \\ \dot{h}_j &= -\frac{h_j}{\tau_r} + \frac{1}{\tau_r \tau_d} \sum_{t_j < t} \delta\left(t - t_{jk}\right) \end{aligned} \tag{9}$$

where τ_r is the synaptic rise time and τ_d is the synaptic decay time. δ is the spike train of *Spike Output* layer (from **Figure 1**), which is filtered through an intermediary exponential filter h_j and in-turn via another exponential filter r_j, which is the final output of the synaptic filter. Following the work of Nicola and Clopath (2017), in which a double exponential synaptic filter was applied to produce smoothed and continuous signals from spike trains, we opted to apply this filter to each output node in the final spiking layer of SNN. This filtered output is transformed into the EEG signal (single-channel EEG for SSVEP and two-channel EEG for MI) according, as shown in the following equation for a single channel c, where * indicates dot product multiplication.

$$EEG_c = \Sigma_j W_{cj} * r_j \tag{10}$$

See **Table 1** for a list of the model parameters.

2.3. Surrogate-Gradient Descent

A spike is emitted when the membrane voltage reaches a predefined threshold ϑ and reset back to resting potential and this behavior mimics Heaviside step activation function (Θ). Due to this non-differentiable nature of spiking neurons, training the synaptic weights is challenging as the traditional gradient

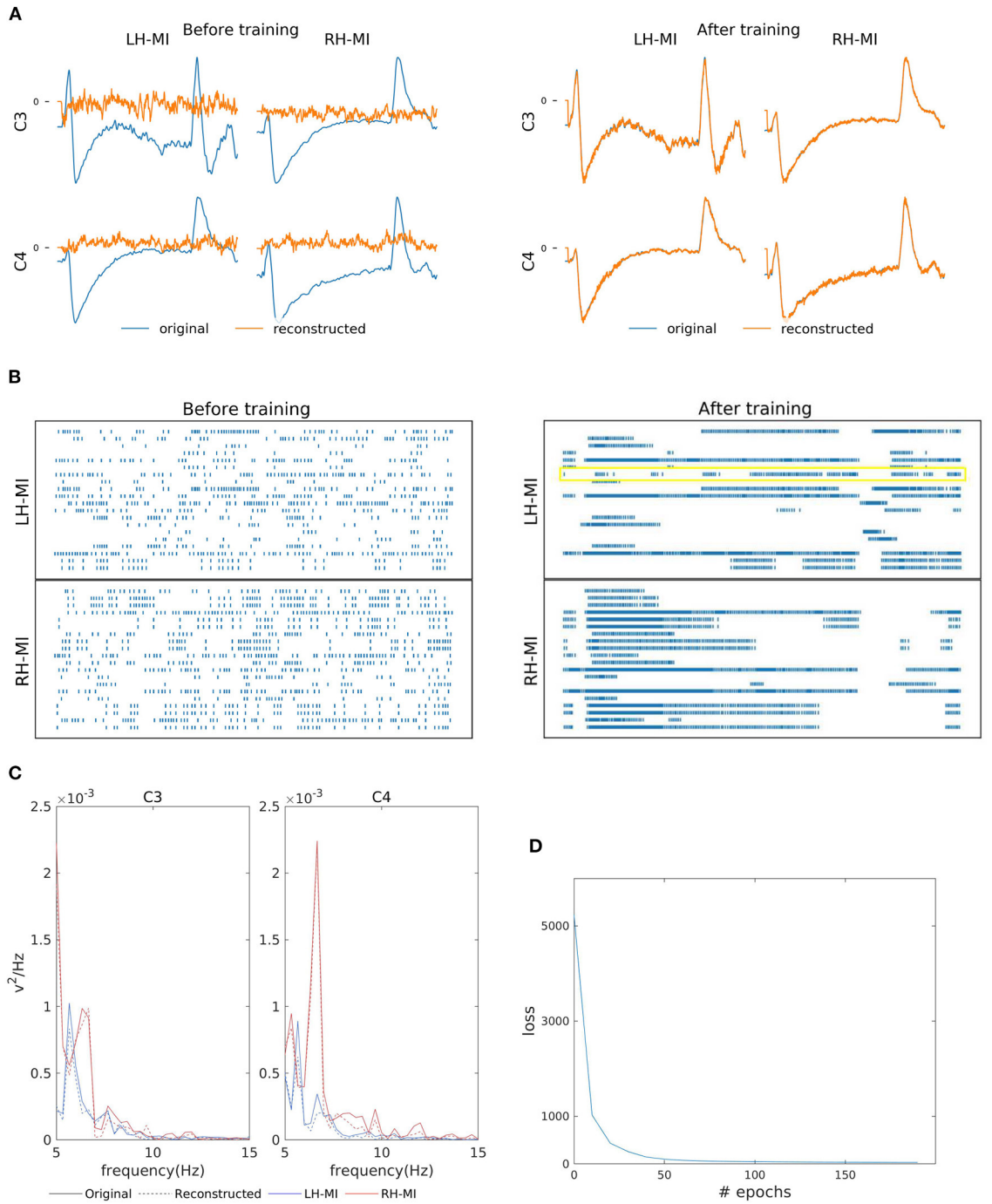


FIGURE 3 | Left-Hand-MI (LH-MI) and Right-Hand MI (RH-MI) EEG signal reconstruction outcomes. **(A)** Illustration of original (blue line) and reconstructed (red line) C3 and C4 EEG signals for LH-MI and RH-MI tasks before and after SNN training. **(B)** Raster plot of 20 randomly selected neuronal spike trains from the hidden layer corresponding to both classes before and after training. The yellow box indicates the effectiveness of the training process in reorganizing the spike trains across layers. However, it should be noted that multiple neuronal spike trains rather than a single neuron contribute to the expected output signal. **(C)** Original (bold line) and reconstructed (dashed line) signal power spectrum at C3 and C4 channels for LH-MI (blue) and RH-MI (red), which appear to be fairly similar. **(D)** The loss value, i.e., the MSE, decreased substantially within a few epochs of the training process.

descent algorithm commonly used for training artificial neural networks (ANNs) is unsuitable because the gradient is zero everywhere except at the event of spike emissions where it is undefined. Additionally, to account for the less extensive hyper-parameter search, in contrast to previous studies such as FORCE, SNNs in this study were trained using a surrogate-gradient

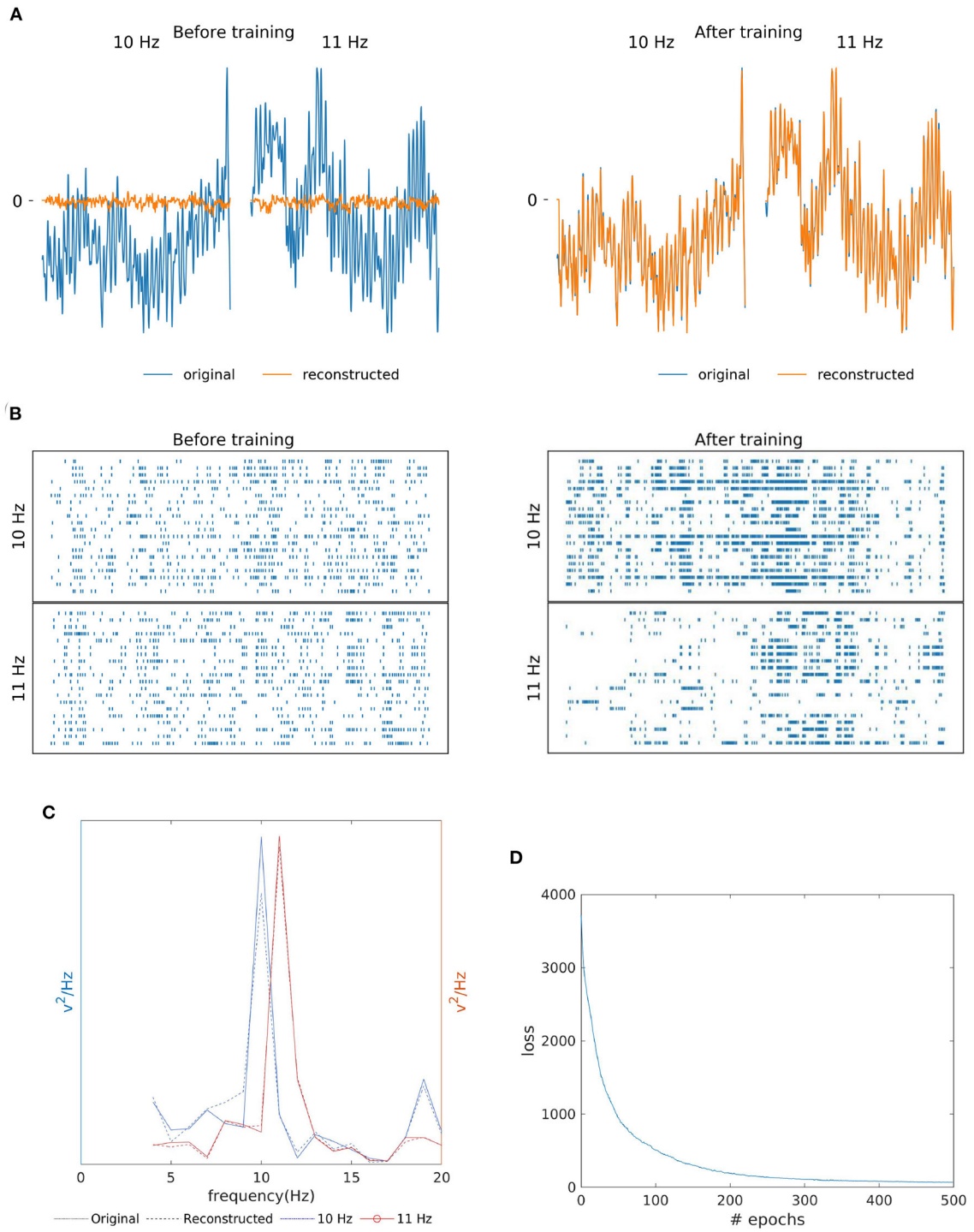


FIGURE 4 | Two class SSVEP signal (10 and 11 Hz) reconstruction outcomes. **(A)** Illustration of original (blue line) and reconstructed (orange line) SSVEP signal of Oz channel before and after SNN training. **(B)** Raster plot of 20 randomly selected neuronal spike trains from the hidden layer corresponding to each class before and after training. The yellow box shows that the training process was effective at reorganizing spike trains. **(C)** Original (bold line) and reconstructed (dashed line) power spectra if signals recorded from the Oz channel are similar, indicating the preservation of the SSVEP biomarker in the reconstructed signal. **(D)** The loss value (MSE) across epochs during the training process.

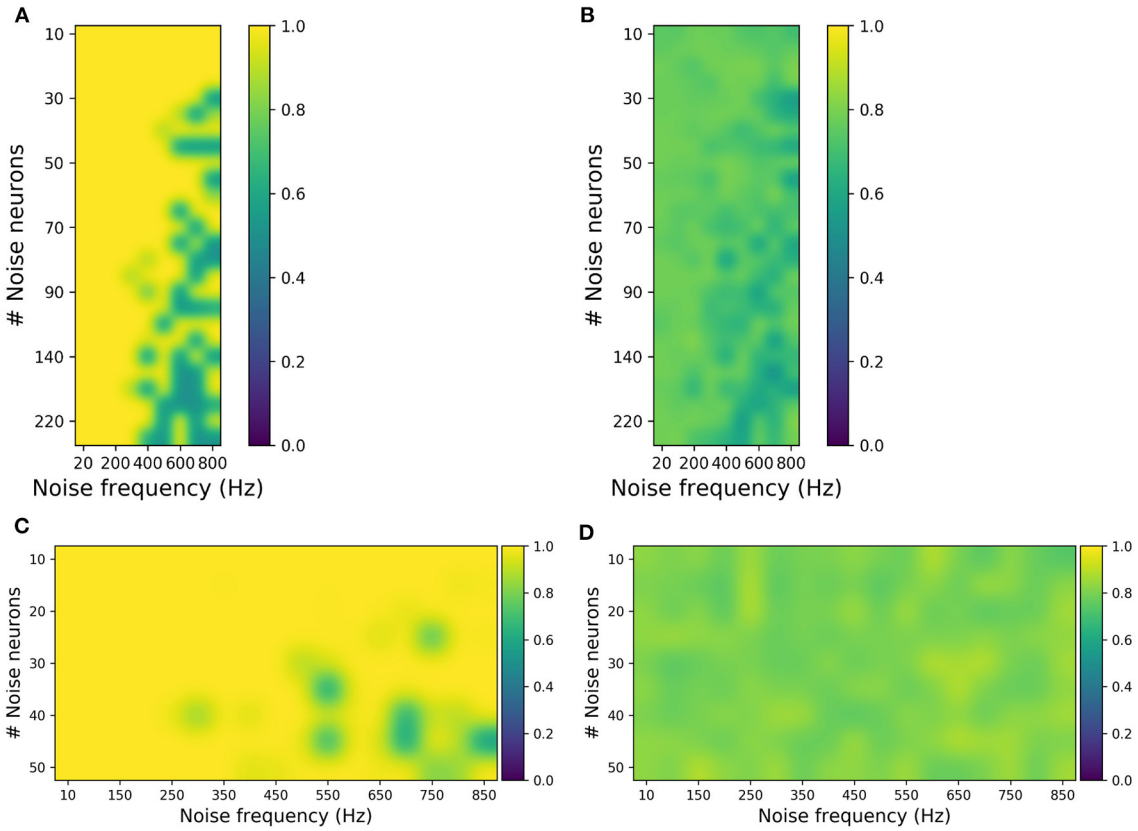


FIGURE 5 | Evaluating synthetic motor imagery (MI) and steady-state visually evoked potential (SSVEP) data generated from different combinations of the number of noise neurons and the firing frequency in the perturbation layer. **(A,B)** MI accuracy obtained by training the classifier with original data and testing with synthetic data (**A**) and vice versa (**B**). **(C,D)** SSVEP accuracy obtained by training classifier with original data and testing with synthetic data **(C)** and vice versa **(D)**. The highest accuracy in **(B)** is ∼80%, while the highest accuracy in **(D)** is 90%, which are close to the accuracies obtained when the original data are used for both training and testing the classifier.

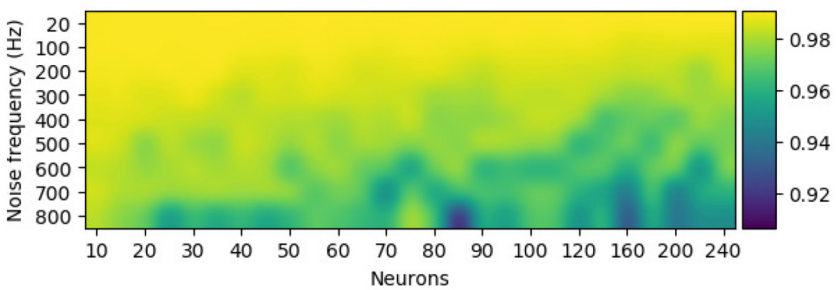


FIGURE 6 | The configuration of the perturbation layer, i.e., the number of neurons and their firing rate, modulates the similarity of the synthetically generated sample to its template for the C4 electroencephalography (EEG) signal. Pearson's correlation between the averaged synthetic samples (100 samples) and the template showcase higher deviation from the template with more neurons and a high firing rate in the perturbation layer.

descent approach (Zenke and Ganguli, 2018; Neftci et al., 2019), which introduces a continuous relaxation on gradient estimation without affecting the forward pass of spike trains. The step nature of spike emission $S_i[n] \propto \Theta(u_i[n]-\vartheta)$ from forward propagation is replaced with $\sigma(u_i[n]-\vartheta)$ during backpropagation, where $\sigma(x) = 1/(1+exp(-x))$. Finally, Adam optimizer (see **Table 1** for properties) was adopted for weight optimization alongside the surrogate gradient.

2.4. EEG Datasets

For MI, the majority of the datasets is from publicly available BCI competition IV resources (Tangermann et al., 2012) and

TABLE 2 | Comparison of classification performance for synthetic motor imagery (MI) assessment when (a) the classifier is trained and tested with original data using 100-fold cross-validation (95% training set and 5% testing set) for baseline accuracy estimation (denoted as *Original*), and (b) the classifier is trained with synthetic data and tested with all of the original data (denoted as *Synthetic*).

Dataset	LDA		SVM		KNN		Gaussian	
	Original	Synthetic	Original	Synthetic	Original	Synthetic	Original	Synthetic
1	80.45	82.50	79.00	**83.13**	81.89	**83.13**	78.44	81.25
2	59.11	63.13	61.56	63.13	54.67	63.13	61.11	**63.75**
3	52.22	**61.25**	51.78	60.63	43.44	60.63	56.56	60.00
4	78.38	67.36	77.63	79.17	74.50	**81.94**	70.13	79.17
5	60.00	65.28	63.00	65.97	**67.75**	66.67	52.75	58.33
6	64.75	65.63	61.88	65.63	56.13	64.38	64.50	**68.75**
7	60.05	**65.83**	62.83	65.00	65.33	63.33	51.83	65.00
8	79.88	**80.71**	78.13	80.71	69.63	80.00	54.00	**80.71**
9	91.00	**93.13**	91.38	93.13	91.00	**93.13**	90.63	91.25
10	71.63	70.63	70.63	69.38	71.50	**72.50**	71.13	**72.50**
11	82.88	**85.00**	83.25	84.38	83.00	84.38	81.00	84.38
12	78.17	77.50	76.17	**78.33**	70.17	**78.33**	69.00	77.50
13	61.63	67.36	59.38	66.67	62.50	67.36	60.88	**70.14**
Avg	70.78	**72.71**	70.51	**73.48**	68.58	**73.76**	66.30	**73.29**

Common spatial pattern (CSP) was used to extract features and was classified with linear discriminant analysis (LDA), support-vector machines (SVM), K-nearest neighbor (KNN), and Gaussian classifier. For all the classifiers, Synthetic has resulted in enhanced performance. LDA resulted in the highest baseline performance of 70.78% and the performance was improved to 72.71% with Synthetic. A Wilcoxon signed-rank test resulted in p-value < 0.01 for all the classifiers. Bold represents the maximum value obtained by a dataset.

few datasets are previously recorded data by our group from Ko et al. (2019). All MI datasets are of 3 s trial duration, sampled to 250 Hz, and contain epochs of left-hand motor imagery and right-hand motor imagery classes. Similarly, publicly available datasets from Wang et al. (2016) were imported for SSVEP analysis. For SSVEP processing, only 2 s epochs were extracted and down-sampled to 250 Hz from the original datasets. Further EEG pre-processing steps for each modality are described in section 3.

2.5. Classification

Since SSVEP is a more effective and high performing BCI modality than motor imagery, this efficacy of data augmentation was explored toward MI classification assessment. Different perturbation combinations (number of noise neurons and their firing rate) in the perturbation layer were used to generate 100 samples (per combination) of artificial MI-EEG data. The artificial MI-EEG data are further assessed for classification performance by (1) comparing the cross-classification between artificial and original data, and (2) correlation analysis to compare the similarity of artificial data with the original. Common spatial pattern (CSP), a widely adopted method for extracting MI-related features (Wang et al., 2006) was used to test the above assessment and the details of the CSP algorithm is described in section 2.5.1. Finally, the features obtained via CSP were classified using linear discriminant analysis (LDA), support-vector machines (SVM) with linear kernel, K-nearest neighbor (K-NN) with k = 5, and Gaussian classifier were adopted (Mishuhina and Jiang, 2018), (Xygonakis et al., 2018).

2.5.1. Common Spatial Pattern

The objective of CSP is to estimate spatial filters for maximal variance to discriminate two sets of data. With E as EEG signal and W as spatial filter, the transformed signal S is given by:

$$S = W^T E \quad \text{or} \quad s(t) = W^T e(t) \tag{11}$$

The criteria for CSP is as follows:

$$\begin{aligned} &\text{maximize} \;\; \text{tr}\, W^T \Sigma_1 \boldsymbol{W} \\ &\text{subject to} \;\; \boldsymbol{W}^T (\Sigma_1 + \Sigma_2)\, \boldsymbol{W} = \boldsymbol{I} \end{aligned} \tag{12}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= \underset{E_n \in \{\text{ class 1}\}}{\text{Exp}} \frac{\boldsymbol{E}_n \boldsymbol{E}_n^T}{\text{tr}\, \boldsymbol{E}_n \boldsymbol{E}_n^T} \\ \boldsymbol{\Sigma}_2 &= \underset{E_n \in \{\text{ class 2}\}}{\text{Exp}} \frac{\boldsymbol{E}_n \boldsymbol{E}_n^T}{\text{tr}\, \boldsymbol{E}_n \boldsymbol{E}_n^T} \end{aligned} \tag{13}$$

3. RESULTS

3.1. SNN for MI

Prominent MI features, such as event-related synchronization (ERS) and event-related desynchronization (ERD), are typically observed over the motor cortex and traced at central EEG electrodes (typically from C3 and C4 channels) in the mu frequency range of 8–13 Hz (Thomas et al., 2009; Shahid et al., 2010). Therefore, in this study, C3 and C4 channel time-series EEG data of two classes: (1) left-hand MI (LH-MI) and (2) right-hand MI (RH-MI) were targeted for data augmentation. The data are band-pass filtered between 0.5 and 20 Hz and sampled at

250 Hz in pre-processing. For extracting the target template, CSP was used to find the best sample from this processed data based on the probability of each sample to be classified accurately. As previously stated, a 10 Hz Poisson spike train matrix is provided as input, one fixed input for each class, and the SNN is trained using surrogate-gradient to produce the C3, C4 channel template. However, during this training process, the perturbation layer (in **Figure 1**) remains inactive.

Figures 3A,B shows the two-channel EEG reconstruction from the training (template) signal before and after the training process and the corresponding reorganization neuron's spiking information from the hidden layer (see the yellow box in **Figure 3B**, indicating that the learning process is effective in reorganizing the spike trains after training). As previously mentioned, for MI reconstruction, C3 and C4 channel data generated by SNN (orange line in **Figure 3A**) were initially random and eventually converged to the EEG template (blue line). We further verified the similarity of reconstructed power spectra to that of the training signals as shown in **Figure 3C**, where the bold lines present the original power spectrum and dashed line represent the reconstructed line for each channel, i.e., C3 and C4 individually for LH-MI (blue line) and RH-MI (red line). **Figure 3D** shows that the loss, i.e., the mean squared error (MSE) between the SNN generated signal and the template signal, converges quickly after a few iterations or epochs. This trained model is later used for generating multiple artificial signals by activating the perturbation layer.

3.2. SNN for SSVEP

SSVEP is the other highly utilized modality for BCI communications. It is detected by the presence of contrasting high power at a frequency similar to the flickering stimulus visually focused on by a user. SSVEP is usually detected from the occipital region, and recordings are typically analyzed from the O1, O2, and Oz electrodes (Zhang et al., 2013). For example, if a user focuses on a 15 Hz flickering stimulus, the EEG signal from the occipital region displays high power at 15 Hz in the frequency domain. Since SSVEP has a high signal-to-noise ratio and is easily traceable with a single electrode, Oz channels data of 10 and 11 Hz SSVEP was considered for SNN-based reconstruction. Contrary to MI, the template signal from SSVEP data is extracted by averaging the available samples. Similar to MI template reconstruction, one fixed 10 Hz Poisson spike train matrix for each class is used as input for reconstructing 10 and 11 Hz Oz's SSVEP signal. Since the targeted frequency lies in the low-frequency range, the dataset was band-pass filtered between 0.5 and 20 Hz and has a sampling rate of 250 Hz.

As shown in **Figures 4A,B**, the similarity of the original and reconstructed signal and hidden layer's spike reorganization before and after SNN training. The SNN training processes were able to successfully re-produce EEG signal (orange line in **Figure 4A)** closely matching the original template (blue line). Additionally, **Figure 4C** shows the resemblance of the reconstructed signal's power spectrum (dashed line) to that of the training signal (bold line). And, for both the 10 Hz signal (blue line) and 11 Hz signal (red line), the reconstruction has the presence of a strong peak (an SSVEP biomarker) at their respective frequencies in the power spectrum. According to **Figure 4D**, similar to MI, the loss (MSE), converged within a few epochs of the training process.

3.3. Synthetic EEG Signal Validation

After successfully training the SNN to reconstruct EEG signal, the neural perturbation layer was activated to generate synthetic EEG data. The number of Poisson neurons (or noise neurons) and the firing rate (or noise frequency) were varied to generate synthetic data in different variance levels. To validate the synthetic data, 100 samples per class were generated for a given combination of neurons and their firing frequency.

Two validation procedures were implemented, *validation-a*: a classifier was trained with the original dataset and tested with the synthetic dataset, and *validation-b*: vice-versa, i.e., a classifier is trained with synthetic data and tested with original data. This validation was performed independently with the data generated by each perturbation combination (see **Figure 5**). The classifiers used for validating artificial MI and SSVEP data include CSP with LDA and multi-layer perceptron (MLP), respectively.

For MI, from **Figure 5A**, it was observed that the synthetic EEG can be classified accurately when the classifier is trained with original data (validation-a). **Figure 5B** shows that although the classification accuracies are not high (for validation-b), the highest accuracy of approximately 80% is similar to the performance when the original dataset is used both for training and testing. And these validations were done independently for each perturbation combination, ranging from 20 to 800 Hz noise frequency and 10–220 noise neurons. This suggests that the generated synthetic data covers the broad variance of the EEG samples. Similarly for SSVEP, **Figures 5C,D** has shown that the synthetic SSVEP passed the validation check with high accuracy. In addition to successful classification, the synthetic data yielded the highest accuracy of approximately 90% when used as a training dataset.

For SSVEP, the perturbation covers a range of 10–50 noise neurons and 10–850 Hz noise frequency. And, similar to MI, SSVEP synthetic validation was done independently for each perturbation combination. Overall, these validation results show that, in addition to reconstructing the EEG signal, the SNN can be used to generate synthetic EEG samples that are in agreement with the original samples.

The configuration of the perturbation layer (i.e., the combination of the number of noise neurons and its firing rate) can be changed to generate data at various scales (or different levels of noise) from the template. Few neurons with low firing rates tend to produce EEG trials that are highly similar to the template, while a perturbation layer with a huge number of neurons and a high firing rate completely distort the synthetic samples of the template. To further check the effect of perturbation combination of artificial data, Pearson's correlation, a metric for estimating the linear correlation between two variables, has been implemented. **Figure 6** shows that for a given EEG channel of a given class (in this case, RH-MI's C4 channel), the correlation between the averaged (100 samples)

synthetic sample and the template is inversely proportional to the number of neurons and the firing rate of the perturbation layer.

3.4. Classification

Since detecting reliable MI features is difficult in comparison to detecting SSVEP, we focused on the classification efficacy of MI-BCI using artificial data. Therefore, for assessing synthetic MI-EEG data, similar to validation-b, a classifier was trained with synthetic data (100 samples per class) and tested with original data. For a baseline comparison, a 100-fold cross-validation process was applied for the original data, in which original data were used for both training and testing. For each fold, 95% of the original data (samples are randomly selected) form a training set and the remaining 5% of the original data were labeled as a test set. This process is repeated for each fold and average accuracy across the fold is estimated to be the baseline accuracy. The synthetic data generated by the perturbation configuration that resulted in the highest accuracy during the validation process (from **Figure 5B**) was used as the final artificial dataset. This process is verified with multiple publicly available datasets (see section 2.4).

Table 2 showcases the accuracy obtained by synthetic data with LDA, SVM, KNN, and Gaussian classifiers. Each classifier was implemented on a dataset basis and in total 13 different datasets were used in this study. When trained with synthetic data, all the classifier was able to successfully predict the labels of original samples and the overall performance was comparable to baseline performance. In particular, LDA resulted in the highest baseline performance (of 70.78%) as commonly stated in literature and the generated synthetic data enhanced the performance to 72.71% with LDA. Considering all the classifiers, a given dataset has shown enhanced performance with synthetic data. Interestingly, for few datasets, the synthetic dataset able to enhance the classification performance up to 9% (for the example dataset 4 with Gaussian classifier). As a cross-validation approach was used to obtain each dataset's baseline accuracy and the distribution of these accuracies across the datasets is unknown, a Wilcoxon signed-rank test was implemented for statistical significance analysis (between performance with synthetic data and original data) and obtained a p-value < 0.01 for all the classifiers.

4. DISCUSSION

Understanding how chaotic neural activity reorganizes to allow humans to perform a wide range of activities, such as walking, running, and speech remains a topic of interest among the neuroscience community. Previous studies have trained models to produce such behavior using different learning methods, such as FORCE, both in artificial neural networks and spiking networks (Sussillo and Abbott, 2009; Nicola and Clopath, 2017; DePasquale et al., 2018). These enforced behaviors are periodic in nature, and training using non-stationary signals such as EEG, is challenging, especially if the algorithm requires a large hyper-parameter search space. Also, EEG signals vary for the same stimulus, i.e., no two EEG samples produced by the human brain are the same, although they retain their prominent biomarker (e.g., ERD/ERS rhythm for MI and high power at flickering stimulus frequency in SSVEP). Since EEG represents an accumulation of potentials of a spiking neuron, a primary focus of this study is to understand how the spiking neural network's activity is organized to form EEG signals and, in turn, generate different samples while retaining the necessary biomarker. Therefore, this study developed a procedure with feedforward SNN and surrogate gradient descent to reconstruct the EEG template in a supervised fashion and used neural perturbation to generate synthetic EEG signals. The proposed model was successfully applied for MI and SSVEP reconstruction, artificial data generation, and validated the synthetic EEG through classification metrics. Additionally, we showed that this network can be used with other BCI modalities, such as P300 (see **Supplementary Figure 1**), and is compatible with other neuron models at a computational cost (see **Supplementary Figures 2**, **3**). To the best of our knowledge, this is the first study to generate synthetic EEG using SNN.

As previously stated, the primary aim of generating synthetic data from a machine learning perspective is to better train a classifier. As shown in **Table 2**, synthetic data can be utilized to enhance BCI performance, especially when the baseline performance of the dataset is low. Although the performance increment was not consistent among all datasets, the proposed approach yielded a significant increase in performance for certain datasets.

In the field of machine learning, the current state-of-the-art model for data augmentation is GAN and its variants. However, to generate artificial data, GAN requires a huge amount of training samples which we find to be counterintuitive. Also, existing EEG based GAN studies normalize the signal from $0\sim1$ or $-1\sim1$ and often fail with EEG data at their original scales. In contrast, the proposed method needs only a few (one or two) samples for generating artificial data and can handle both normalized and non-normalized EEG signals. To this end, we compared recently published generative models and GAN variants, such as VAE, W-GAN, and DCGAN, and class-conditioned W-GAN (cc-WGAN) (Aznan et al., 2019; Panwar et al., 2020), with SNN for MI data generation and validation tests. The cc-WGAN model was used to generate MI data for few datasets. And previous studies usually mix the samples of original and artificial for classification enhancement often fail to perform *validation-b*. Appropriate synthetic data apart from enhancing the classification performance should also pass this validation check (see section 3.3). According to **Supplementary Table 1**, we observed that the GAN-generated data failed in this validation check with below chance level. The reason for this could be of multitude such as the random noise in the GAN-generated data improved the performance by chance, or limited amount of data (the GAN was trained independently for each dataset containing approximately 180 trials) failed to capture MI's biomarker. Contrarily, the proposed SNN approach passed this validation check (**Supplementary Table 1**) and delivered close to baseline or improved performance. Another major reason for this could be because, since SNN generates data based on

a template signal, the artificial trials with varying distribution retrained MI-biomarker, and it in-turn able to make the classifier extract better features. Since GAN accounts for the whole dataset for training, the artifacts and noise present in the data could have lead to the generation of improper samples. This hypothesis could be explored in the future to fine-tune GAN's and SNN's training samples.

In agreement with previous studies, we interestingly observed a reorganization of neural activity (i.e., spike train) concerning the output signal. For example, the SSVEP signal at a defined frequency is slightly oscillatory (like a sine or cosine wave), which is reflected in the periodic spiking activity of certain neurons in the output layer (see **Figure 4B**). We further tested the proposed model for robustness by randomly clipping synapses and found that the overall signal trend was maintained despite the 30% reduction in synapses (see **Supplementary Figure 4**). According to Dale's law, a neuron is only either excitatory or inhibitory (Capano et al., 2015). However, this constraint was omitted from this study for simplicity, although it would be ideal to enforce the non-stationary behavior, such as in EEG, to spiking neural networks to maintain Dale's law in the future. Additionally, the current study, being focused on MI and SSVEP data, considered only a few channels. However, other minor BCI-related concepts such as fatigue, attention, and other physiological factors distribute their feature across a range of EEG channels. Therefore, another future direction of this study could be to test the limits of the model in terms of the number of channels and categories.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: http://www.bbci.de/competition/iv/#dataset2a; http://www.bbci.de/competition/iv/#dataset2b; http://bci.med.tsinghua.edu.cn/download.html.

AUTHOR CONTRIBUTIONS

SS and C-TL conceptualized the concept and wrote the manuscript. SS wrote the code and performed the simulations. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Australian Research Council (ARC) under discovery grant DP180100670 and DP180100656, the Australia Defence Innovation Hub under Contract No. P18-650825, and the US Office of Naval Research Global under Cooperative Agreement Number ONRG-NICOP-N62909-19-1-2058.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.651762/full#supplementary-material

REFERENCES

Alwasiti, H. H., Aris, I., and Jantan, A. (2010). Brain computer interface design and applications: challenges and future. *World Appl. Sci. J.* 11, 819–825.

Aznan, N. K. N., Atapour-Abarghouei, A., Bonner, S., Connolly, J., Moubayed, N. A., and Breckon, T. (2019). Simulating brain signals: creating synthetic EEG data via neural-based generative models for improved SSVEP classification. *arXiv preprint arXiv:1901.07429*. doi: 10.1109/IJCNN.2019.8852227

Capano, V., Herrmann, H. J., and De Arcangelis, L. (2015). Optimal percentage of inhibitory synapses in multi-task learning. *Sci. Rep.* 5:9895. doi: 10.1038/srep09895

Chiarelli, A. M., Croce, P., Merla, A., and Zappasodi, F. (2018). Deep learning for hybrid EEG-fNIRS brain-computer interface: application to motor imagery classification. *J. Neural Eng.* 15:036028. doi: 10.1088/1741-2552/aaaf82

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., et al. (2012). Neural population dynamics during reaching. *Nature* 487:51. doi: 10.1038/nature11129

Cincotti, F., Mattia, D., Aloise, F., Bufalari, S., Astolfi, L., Fallani, F. D. V., et al. (2008). High-resolution EEG techniques for brain-computer interface applications. *J. Neurosci. Methods* 167, 31–42. doi: 10.1016/j.jneumeth.2007.06.031

DePasquale, B., Cueva, C. J., Rajan, K., Escola, G. S., and Abbott, L. (2018). full-FORCE: a target-based method for training recurrent networks. *PLoS One* 13:e0191527. doi: 10.1371/journal.pone.0191527

Dinarés-Ferran, J., Ortner, R., Guger, C., and Solé-Casals, J. (2018). A new method to generate artificial frames using the empirical mode decomposition for an eeg-based motor imagery BCI. *Front. Neurosci.* 12:308. doi: 10.3389/fnins.2018.00308

Donoghue, J. P. (2002). Connecting cortex to machines: recent advances in brain interfaces. *Nat. Neurosci.* 5:1085. doi: 10.1038/nn947

Fazli, S., Mehnert, J., Steinbrink, J., and Blankertz, B. (2012). "Using NIRS as a predictor for EEG-based BCI performance," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Berlin), 4911–4914. doi: 10.1109/EMBC.2012.6347095

Gilra, A., and Gerstner, W. (2017). Predicting non-linear dynamics by stable local learning in a recurrent spiking neural network. *Elife* 6:e28295. doi: 10.7554/eLife.28295

Hartmann, K. G., Schirrmeister, R. T., and Ball, T. (2018). EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *arXiv preprint arXiv:1806.01875*.

Ingrosso, A., and Abbott, L. (2019). Training dynamically balanced excitatory-inhibitory networks. *PLoS ONE* 14:e220547. doi: 10.1371/journal.pone.0220547

Jin, J., Sellers, E. W., Zhou, S., Zhang, Y., Wang, X., and Cichocki, A. (2015). A p300 brain-computer interface based on a modification of the mismatch negativity paradigm. *Int. J. Neural Syst.* 25:1550011. doi: 10.1142/S0129065715500112

Ko, L.-W., Lu, Y.-C., Bustince, H., Chang, Y.-C., Chang, Y., Ferandez, J., et al. (2019). Multimodal fuzzy fusion for enhancing the motor-imagery-based brain computer interface. *IEEE Comput. Intell. Mag.* 14, 96–106. doi: 10.1109/MCI.2018.2881647

Lotte, F. (2011). "Generating artificial EEG signals to reduce BCI calibration time," in *5th International Brain-Computer Interface Workshop* (Bordeaux), 176–179.

Mainen, Z. F., and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science* 268, 1503–1506. doi: 10.1126/science.7770778

Marchesotti, S., Bassolino, M., Serino, A., Bleuler, H., and Blanke, O. (2016). Quantifying the role of motor imagery in brain-machine interfaces. *Sci. Rep.* 6:24076. doi: 10.1038/srep24076

Mishuhina, V., and Jiang, X. (2018). Feature weighting and regularization of common spatial patterns in EEG-based motor imagery BCI. *IEEE Signal Process. Lett.* 25, 783–787. doi: 10.1109/LSP.2018.2823683

Moore, M. M. (2003). Real-world applications for brain-computer interface technology. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 162–165. doi: 10.1109/TNSRE.2003.814433

Nakanishi, M., Wang, Y., Wang, Y.-T., Mitsukura, Y., and Jung, T.-P. (2014). A high-speed brain speller using steady-state visual evoked potentials. *Int. J. Neural Syst.* 24:1450019. doi: 10.1142/S0129065714500191

Neftci, E. O., Mostafa, H., and Zenke, F. (2019). Surrogate gradient learning in spiking neural networks. *arXiv preprint arXiv:1901.09948*.

Nicola, W., and Clopath, C. (2017). Supervised learning in spiking neural networks with force training. *Nat. commun.* 8:2208. doi: 10.1038/s41467-017-01827-3

Panwar, S., Rad, P., Jung, T.-P., and Huang, Y. (2020). Modeling EEG data distribution with a wasserstein generative adversarial network to predict rsvp events. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1720–1730. doi: 10.1109/TNSRE.2020.3006180

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* 51, 1034–1043. doi: 10.1109/TBME.2004.827072

Schwemmer, M. A., Skomrock, N. D., Sederberg, P. B., Ting, J. E., Sharma, G., Bockbrader, M. A., et al. (2018). Meeting brain-computer interface user performance expectations using a deep neural network decoding framework. *Nat. Med.* 24:1669. doi: 10.1038/s41591-018-0171-y

Shahid, S., Sinha, R. K., and Prasad, G. (2010). Mu and beta rhythm modulations in motor imagery related post-stroke EEG: a study under BCI framework for post-stroke rehabilitation. *BMC Neurosci.* 11:P127. doi: 10.1186/1471-2202-11-S1-P127

Sussillo, D., and Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63, 544–557. doi: 10.1016/j.neuron.2009.07.018

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6:55. doi: 10.3389/fnins.2012.00055

Thomas, K. P., Guan, C., Lau, C. T., Vinod, A. P., and Ang, K. K. (2009). A new discriminative common spatial pattern method for motor imagery brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 56, 2730–2733. doi: 10.1109/TBME.2009.2026181

Wang, Y., Chen, X., Gao, X., and Gao, S. (2016). A benchmark dataset for SSVEP-based brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1746–1752. doi: 10.1109/TNSRE.2016.2627556

Wang, Y., Gao, S., and Gao, X. (2006). "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (Beijing), 5392–5395. doi: 10.1109/IEMBS.2005.1615701

Xygonakis, I., Athanasiou, A., Pandria, N., Kugiumtzis, D., and Bamidis, P. D. (2018). Decoding motor imagery through common spatial pattern filters at the EEG source space. *Comput. Intell. Neurosci.* 2018:7957408. doi: 10.1155/2018/7957408

Zenke, F., and Ganguli, S. (2018). Superspike: Supervised learning in multilayer spiking neural networks. *Neural Comput.* 30, 1514–1541. doi: 10.1162/neco_a_01086

Zhang, J., Yan, C., and Gong, X. (2017). "Deep convolutional neural network for decoding motor imagery based brain computer interface," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (Shanghai), 1–5. doi: 10.1109/ICSPCC.2017.8242581

Zhang, Y., Zhou, G., Jin, J., Wang, M., Wang, X., and Cichocki, A. (2013). L1-regularized multiway canonical correlation analysis for SSVEP-based BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21, 887–896. doi: 10.1109/TNSRE.2013.2279680

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fnins.2021.646013

Bidirectional Mapping of Brain MRI and PET With 3D Reversible GAN for the Diagnosis of Alzheimer's Disease

Wanyun Lin[1,2], Weiming Lin[3], Gang Chen[4,5], Hejun Zhang[4,5], Qinquan Gao[1,6], Yechong Huang[7], Tong Tong[1,6], Min Du[1,2]* and the Alzheimer's Disease Neuroimaging Initiative[†]*

[1] College of Physics and Information Engineering, Fuzhou University, Fuzhou, China, [2] Fujian Provincial Key Laboratory of Medical Instrument and Pharmaceutical Technology, Fuzhou University, Fuzhou, China, [3] School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen, China, [4] Department of Pathology, Fujian Cancer Hospital & Fujian Medical University Cancer Hospital, Fuzhou, China, [5] Fujian Provincial Key Laboratory of Translational Cancer Medicine, Fuzhou, China, [6] Imperial Vision Technology, Fuzhou, China, [7] School of Data Science, Fudan University, Shanghai, China

Combining multi-modality data for brain disease diagnosis such as Alzheimer's disease (AD) commonly leads to improved performance than those using a single modality. However, it is still challenging to train a multi-modality model since it is difficult in clinical practice to obtain complete data that includes all modality data. Generally speaking, it is difficult to obtain both magnetic resonance images (MRI) and positron emission tomography (PET) images of a single patient. PET is expensive and requires the injection of radioactive substances into the patient's body, while MR images are cheaper, safer, and more widely used in practice. Discarding samples without PET data is a common method in previous studies, but the reduction in the number of samples will result in a decrease in model performance. To take advantage of multi-modal complementary information, we first adopt the Reversible Generative Adversarial Network (RevGAN) model to reconstruct the missing data. After that, a 3D convolutional neural network (CNN) classification model with multi-modality input was proposed to perform AD diagnosis. We have evaluated our method on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, and compared the performance of the proposed method with those using state-of-the-art methods. The experimental results show that the structural and functional information of brain tissue can be mapped well and that the image synthesized by our method is close to the real image. In addition, the use of synthetic data is beneficial for the diagnosis and prediction of Alzheimer's disease, demonstrating the effectiveness of the proposed framework.

Keywords: Alzheimer's disease, multi-modality, image synthesis, 3D CNN, reversible GAN

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Young Don Son, Gachon University Gil Hospital Center, South Korea
Yi Su, Banner Alzheimer's Institute, United States

***Correspondence:**
Tong Tong
ttraveltong@gmail.com
Min Du
dm_dj90@163.com

[†]Data used in the preparation of this paper were obtained from the ADNI database (http://adni.loni.usc.edu/). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at the website (http://adni.loni.usc.edu/wpcontent/uploads/how\protect_to\protect_apply/ADNI\protect_Acknowledgement\protect_List.pdf)

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *24 December 2020*
Accepted: *11 March 2021*
Published: *09 April 2021*

Citation:
Lin W, Lin W, Chen G, Zhang H, Gao Q, Huang Y, Tong T, Du M and the Alzheimer's Disease Neuroimaging Initiative (2021) Bidirectional Mapping of Brain MRI and PET With 3D Reversible GAN for the Diagnosis of Alzheimer's Disease. Front. Neurosci. 15:646013. doi: 10.3389/fnins.2021.646013

INTRODUCTION

Alzheimer's disease (AD) is a common neurodegenerative disease and there is no cure for AD so far. Relevant researches show that AD accounts for approximately 60–70% of patients with dementia. Different modalities of neuroimaging can reflect disease changes of AD from different perspectives. Recent studies have shown that MR images and PET images contain complementary information in AD diagnosis (Liu et al., 2017). However, it is difficult to obtain complete modality data for all

individuals. Subjects may lack a specific modality due to the high cost and the usage of radioactive tracers, which will increase lifetime cancer risk. In clinical practice, subjects are more willing to accept MRI scans than PET scans due to price and safety considerations. Therefore, collecting a large number of paired data in AD research is a challenge.

In Calhoun and Sui (2016) study, they directly discard samples with incomplete modalities. This reduces the number of samples available for training. The lack of training data may lead to the overfitting problem, thus resulting in poor diagnosis performance. In the past few years, studies on medical image synthesis tasks have been performed. They use algorithms to estimate missing data instead of simply discarding incomplete samples. For example, Li et al. (2014) applied a 3D convolutional neural network (CNN) model to predict PET images from MR images. Moreover, Pan et al. (2018) proposed a 3D-cGAN model to estimate the corresponding PET data based on MRI data. The synthetic data is used for AD classification (Pan et al., 2019) developed a deep learning model called disease-image specific neural network (DSNN), which can simultaneously perform image synthesis and disease classification tasks. Additionally, Nie et al. (2017) synthesized CT images from corresponding MR images using a cascaded 3D full-convolution network (Zhao et al., 2018) adopted Deep-supGAN learning maps between 3D MR data and CT image. Similarly, Bi et al. (2017) synthesized high-resolution PET images from paired CT images. BenTaieb and Hamarneh (2017) attempted to solve the staining problem by training cGAN and task-specific networks (segmentation or classification models) (Wei et al., 2019) predicted PET-derived demyelination from multiparametric MRI. The above studies demonstrate that GAN is a powerful technique for data simulation and expansion in segmentation or classification tasks. However, there is still much room to improve the performance in many medical image synthesis tasks. Some state-of-the-art methods are one-way image synthesis (for example, generating PET from MRI, generating CT from MRI, etc.), which cannot maximize the expansion of missing datasets. Other methods have used very complex preprocessing steps, resulting in high computational costs and difficult in reproducing their results.

In this paper, we imputed the missing data through the 3D Reversible Generative Adversarial Network (RevGAN), and compared the advantages and disadvantages of using synthetic full image and synthetic ROI image. In addition, we also use the synthesized data for AD classification prediction. The main contributions of this study are as follows: First, the reversible structure was utilized to yield better reconstruction ability in the image synthesis experiment. We have improved the generator part, and only one generator was needed to realize bidirectional image synthesis in the proposed method rather than the two generators that were used in the methods of Pan et al. (2018). By adopting the generator together with the stability of reversible architecture, this allows us to train deeper networks using only modest computational resources. Second, by comparing the synthesis experiment of full image and ROI image, we can find that the structural and functional information of brain tissue can be mapped well, but it is difficult to map the structure information such as the skull of the MR image from the PET image. Third, in the classification experiment using full image, we can find that the classification model is mainly based on the brain tissue area in the neuroimaging and is not sensitive to the skull and other structures by comparing the experiment using real MRI and synthetic MRI. Fourth, by comparing the missing data and the use of synthetic data, our proposed image synthesis method is not only of high image quality but also contains disease information about AD, which can be beneficial for the auxiliary diagnosis of diseases. In addition, we also designed a multi-modal 3D CNN model that performed well on the data we used. In the following section, we will first introduce the dataset that we used for evaluation in section "Materials." We introduce the preprocessing steps and the details of the proposed method in section "Methods." The experiments and results are shown in section 4 "Experiments and Results," which mainly includes three parts: experimental setup, image reconstruction and the impact of using synthetic data on Alzheimer's diagnosis and prediction. The discussion and the conclusion of our work are described in sections "Discussion" and "Conclusion," respectively.

MATERIALS

Datasets

The data used in this study comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which publicly provides a series of test subjects' MRI, PET, other biomarkers and related diagnostic information, providing researchers with A set of standard research data used to study the pathogenesis of Alzheimer's disease. The data provided by it contains four sets of sub-libraries, namely ADNI-1, ADNI-2, ADNI-3, and ADNI GO. These four stages contain data from subjects in three categories: cognitively unimpaired (labeled as CN), mild cognitive impairment (MCI) and AD. In the problem of predicting the conversion of MCI to AD, it is necessary to review the condition of the MCI subjects so as to keep up with the progress of the subject's condition. Data will be collected again six months, twelve months, eighteen months, twenty-four months, and thirty-six months after the baseline data is collected. Generally speaking, the time standard for judging whether MCI is converted is thirty-six months. Subjects who converted from MCI to AD within 36 months belonged to developmental mild cognitive impairment (pMCI), and vice versa were classified as stable mild cognitive impairment (sMCI). Although there is no cure for AD so far, the ADNI database has greatly facilitated research on AD by researchers. **Table 1** summarizes the details of the data we used.

TABLE 1 | Summary of the studied subjects and images from the dataset.

Class	Subject Number	Image Number
AD	362	647
CN	308	707
pMCI	183	326
sMCI	233	396

METHODS

There are three major steps in the proposed framework: (i) Data preprocessing step; (ii) Missing data completion using 3D Reversible GAN; (iii) Disease classification using 3D convolutional neural networks. The overall framework of the proposed approach is shown in **Figure 1**, and the above steps are introduced in the following subsections, respectively.

In our research, we assume M^k and P^k are the MRI data and PET data for the kth subject, respectively. The diagnosis result of the model can be expressed as:

$$\hat{y^k} = F(M^k, P^k), \tag{1}$$

where $\hat{y^k}$ is the predicted label (such as sMCI/pMCI) for the kth subject. If the kth subject does not have PET data (missing P^k), our method will predict a virtual $\hat{P^k}$ with M^k by their underlying relevance. The diagnosis result of the model can be expressed as:

$$\hat{y^k} = F(M^k, P^k) \approx F(M^k, G(M^k)) \tag{2}$$

G and F are both mapping functions. It can be seen from the above formula that there are two major tasks in our framework, (i) learning a mapping function G for completing the missing data, which is described in section "Data Reconstruction Using 3D RevGAN." (ii) learning a classification model F for AD diagnosis and prediction, which is then introduced in section "Classification With End-to-End CNNs."

Data Preprocessing

All images were preprocessed based on the following steps as described in Huang et al. (2019). The **Figure 2** shows the data processing steps. The MR brain images were preprocessed by the standard ADNI pipeline as described in Jack et al. (2008), which includes post-acquisition correction of gradient warping, B1 non-uniformity correction, intensity non-uniformity correction and phantom-based scaling correction. After that, we performed ITK N4 Bias correction on the MR image. Based on zxwtools, MR images were resampled to a 1 mm isotropic world coordinate system and cropped to a size of 221 × 257 × 221. Since the pixel value range of the MR image is long-tailed, we set voxels beyond 1024 as 1024. The pixel values range of the MR images were unified into [0,1024], using the formula of 1024×(input-min)/(max-min). For PET data, each patient may have multiple images, and each Image contains multiple.nii files (30 min to 60 min from taking the medicine, taking pictures at regular intervals). After PET undergoes image quality control (removal of outliers, such as images with 0 mean and variance), we normalize the PET image to [0, 1024] according to the formula 1024×(input-min)/(max-min). After that, we averaged all PET images of the same patient under the same radiography. Finally, PET images were registered to the corresponding MR based on zxhreg through rigid registration. After the registration, PET images were resampled to a 1 mm isotropic spacing.

In this study, the hippocampus region was selected as the ROI. We obtain the center of the ROI as follows: First, for each subject, the hippocampus was segmented using the MALP-EM (Heckemann et al., 2015; Ledig et al., 2015). After getting the segmentation result, we directly calculated the center of the hippocampus (the segmentation result already shows this value). After that, we randomly selected MR data as a template and then registered other MR images to the template through affine registration to obtain a parameter matrix of affine transformation. After obtaining the affine matrix, we map the hippocampus center point on the template to each MR image through the inverse transformation of the affine transformation. Since MR and PET have been registered before, the ROI center of the obtained MR can also be used as the ROI center of PET. After determining the ROI center of the image, we cropped a 96 × 96 × 48 voxel hippocampus region from the image.

Before the data enters the model, the pixel values range of the data is unified into [0,1] using the formular of (input-min)/(max-min). The full image and the corresponding ROI image are shown in **Figure 3**.

Data Reconstruction Using 3D RevGAN

In recent years, a lot of studies have widely used GAN (Goodfellow et al., 2014; Yu et al., 2017) in the field of medical image generation. Problems such as data shortage and class imbalance can be solved by GAN (Frid-Adar et al., 2018), and help to understand the data distribution and its potential

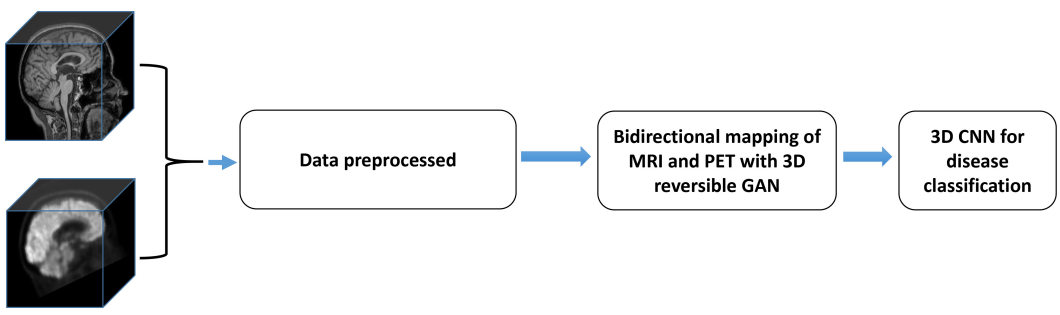


FIGURE 1 | The overall flow chart of the proposed method. It includes three parts, and the data is preprocessed first. RevGAN was then used to synthesize MRI and PET and to evaluate the image quality. Finally, we verify the impact of the synthesized images on disease classification.

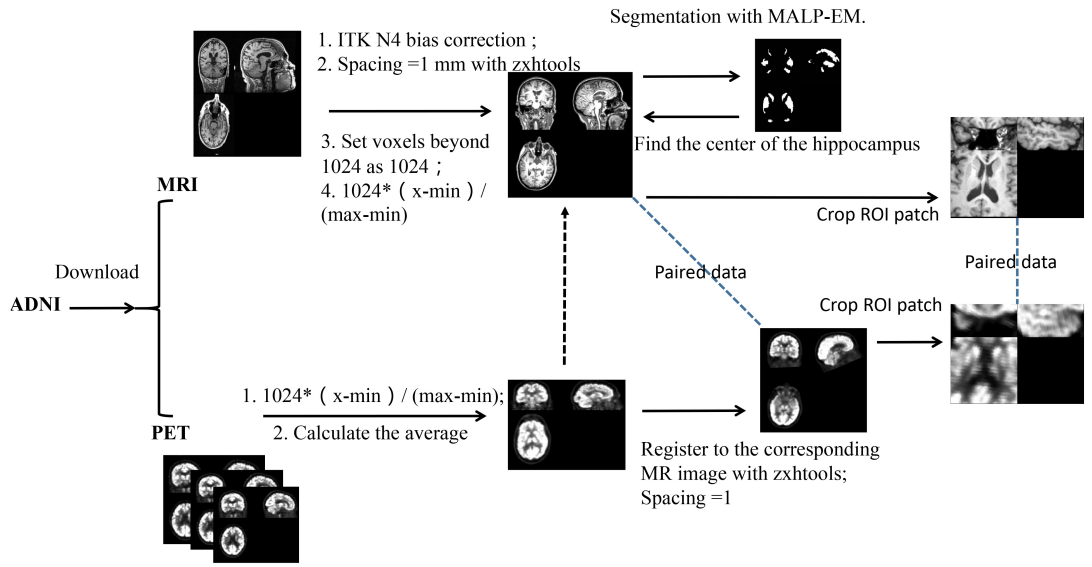


FIGURE 2 | The pipeline of data preprocessing.

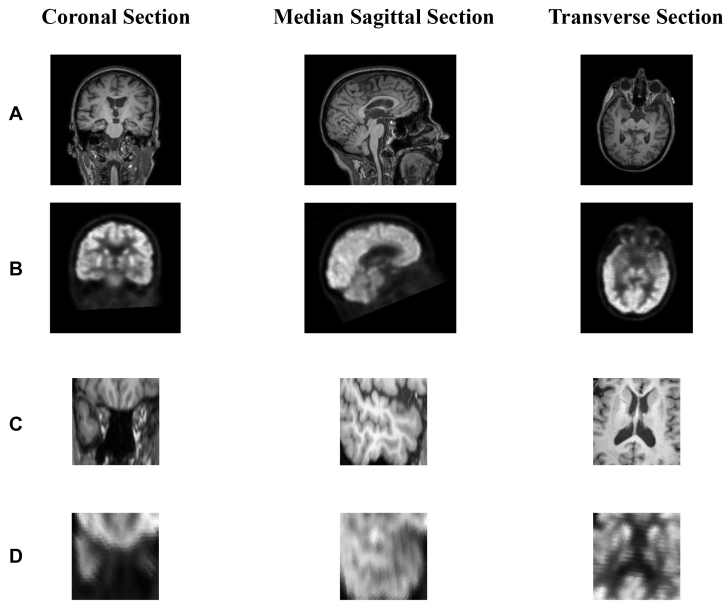


FIGURE 3 | (A,B) show the coronal section, median sagittal section and transverse section of MR image and PET image, respectively. **(B)** is registered to the corresponding **(A)** through rigid registration. The dimensions of **(A,B)** are the same as 221 × 257 × 221. **(C)** is the area near the hippocampus of the MR image and **(C)** is obtained from **(A)**. Since **(A,B)** are registered, we can obtain **(D)** corresponding to **(C)** from **(B)**. In short, **(A,B)** or **(C,D)** represents the same area in the brain.

structure. In this study, we propose an image generation model based on the RevGAN (van der Ouderaa and Worrall, 2019). The proposed framework of our 3D RevGAN model is illustrated in **Figure 4**, which includes a reversible generator (G), and also two adversarial discriminators (D1, D2). The generator consists of three sequential parts (encoder, invertible core, decoder). The encoder part is constructed by a 3 × 3 × 3 convolutional layer for extracting the knowledge of images. The instance normalization layer and ReLU activation function follow the convolutional layer. The encoder maps the image into a higher dimensionality space. The invertible core C and its inverse C^{-1} are composed of many 3D reversible blocks, which can transfer knowledge from the original domain to the target domain. In our model, we use invertible residual layer, as used in Gomez et al. (2017), using additive coupling (Dinh et al., 2014) as a reversible block. In this work, two reversible blocks were used. Each reversible block is composed of a subnetwork R1 and a subnetwork R2 as shown in **Figure 5**. R1 and R2 are not reversible. This layer is very

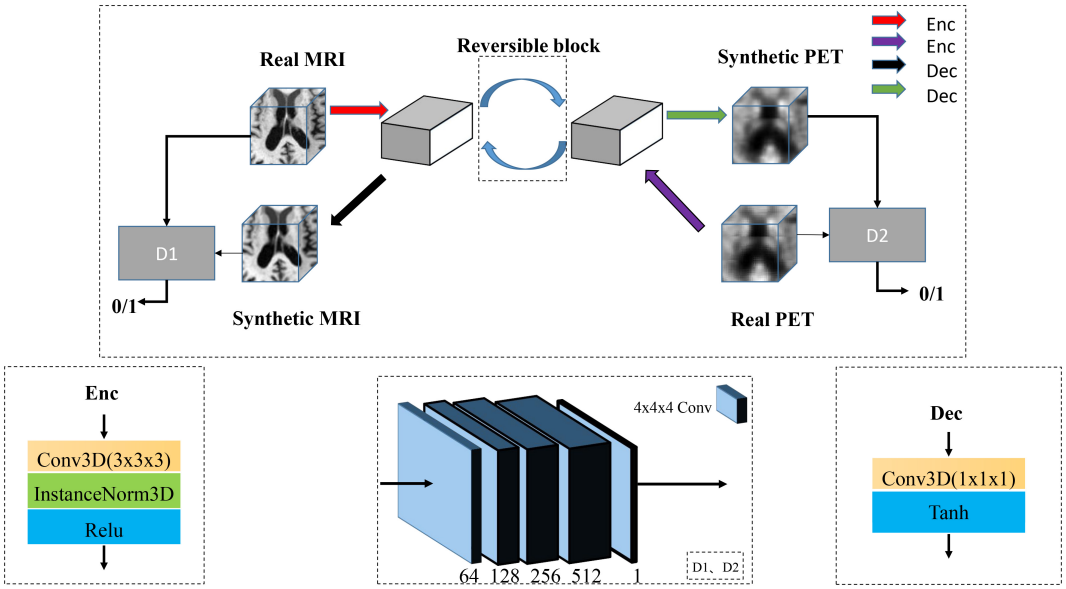


FIGURE 4 | Illustration of our proposed image synthesis method using 3D-RevGAN for learning the mappings between MRI and PET. Encoders Enc (red) and Enc (purple) lift/encode from the image space into the feature spaces. Decoders Dec (black) and Dec (green) project/decode back to the image space. Compared with CycleGAN using two generators, RevGAN only uses one generator to complete the bidirectional mapping from MRI to PET and PET to MRI.

flexible and easy to implement. The detailed structures of R1 and R2 are shown in **Figure 5**, which consist of two convolutional layers, one normalization layer and one non-linear activation layer. The network structure is based on SRCNN (Dong et al., 2014). R1 and R2 in **Figure 5** satisfy the following relationship:

$$y_1 = x_1 + R_1(x_2) \quad x_1 = y_1 - R_1(x_2), \tag{3}$$

$$y_2 = x_2 + R_2(y_1) \quad x_2 = y_2 - R_2(y_1), \tag{4}$$

Since core C and its inverse C^{-1} share parameters, C^{-1} will also be trained when training C, and vice versa. Finally, the decoder part is constructed by a $1 \times 1 \times 1$ convolutional layer for constructing the images in the target domain and is directly followed by a tanh activation function. Decoder projects the image back into the low dimensional image space. For the discriminator, we adopt the PatchGAN architecture as proposed in Zhu et al. (2017). It inputs a pair of real image (such as P^k) and synthetic image [$G(M^k)$]. We combine the adversarial loss (L_{GAN}) loss with the cycle consistency loss (L_{cyc}), so the total loss function used in our model is as follows:

$$L_{RevGAN} = L_{GAN}(G, D1) + L_{GAN}(G^{-1}, D2) + \lambda L_{cyc}(G, G^{-1}) \tag{5}$$

The weighting factor λ (called lambda) is used to control the weight of the cyclic consistency loss in the total loss.

Classification With End-to-End CNNs

After the missing data of multi-modality have been completed using the above steps, the diagnosis and prediction of AD are then performed using a classifier. In some previous studies, the hippocampus regions have been chosen as the ROI for the diagnosis of AD (Dickerson et al., 2001; Schuff et al., 2009; Salvatore et al., 2015). The ROI selection is an important step in the proposed method in clinical practice. The hippocampus area is of great significance in the diagnosis of AD, and has been used as a biomarker in diagnostic criteria for Alzheimer's disease. The hippocampal atrophy is highly related to AD and can be reflected in MR and PET images. The brain atrophy of AD patients affects both the structural change and the metabolic function change at the same time. Lots of previous studies have used the hippocampus area as ROI for the prediction and diagnosis of AD. In contrast, if a full image is used, although more disease information can be used, the information of a large number of irrelevant regions in the full image coexist and may hamper the accuracy of the prediction model. The use of the hippocampus as an ROI is not only essential to improve the diagnostic accuracy, but also can reduce the computational burden and reduce the data redundancy. For a fair comparison, we also performed experiments using full images and use scipy.ndimage.zoom to scale the registered full image size from (221, 257, 221) to (110, 128, 110).

CNN has been well applied in the classification and prediction of brain disease. Briefly, CNN extracts feature through convolutional layers and reduces the network parameters through weight sharing and pooling of convolution. Finally, the classification task is completed through the fully connected layer. The subsequent development of many advanced networks such as AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), ResNet (He et al., 2016)

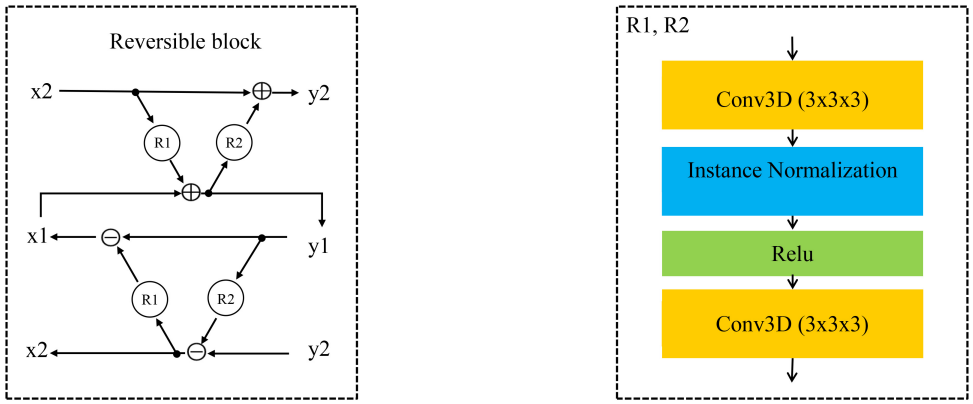


FIGURE 5 | The reversible block is composed of R1 and R2.

and so on. In our experiment, in order to evaluate the performance of our proposed 3D CNN, we chose the classic model of VGG and ResNet as the baseline model for comparisons.

Although MR and PET share most information of AD disease changes, it should be mentioned that these two modalities also have complementary information (Fan et al., 2008; Wee et al., 2014; Calhoun and Sui, 2016), which can be beneficial for further improvement of AD diagnosis. The process of synthesized PET can generate complementary information. Although this information comes from MRI, the classification model cannot use this information directly. Through our synthesis method, these hidden information are displayed, so that the classification network can obtain more information. For example, elements (5, 6, 7) are in set A, (15, 16, 17) are in set B, and B = A +10. As long as we know the relationship between AB, we can infer each other. However, in our case A does not contain B, and B does not contain A. In this study, the relationship between MRI and PET is a more complex non-linear complementary relationship, and our model is needed to fit the relationship between them. The synthetic network is trained through a large number of pairs of other people's MRI and PET relationships. During the training process, the synthetic network learns the method of converting the information in the MRI into the corresponding PET information.

We proposed a 3D CNN model based on ROI crop to learn a classifier and to fuse different features from both MRI and PET. It is worth noting that the general classification model uses batch norm, while we use instance norm (Ulyanov et al., 2016). The output of each convolutional layer is down-sampled by the max-pooling operations. In the 1st, 2nd, 3rd, and 4th layers, the size of the convolutional kernel is set to $1 \times 1 \times 1$, $5 \times 5 \times 5$, $9 \times 9 \times 9$, $5 \times 5 \times 5$, respectively. The number of channels is 4 for the 1st layer, 32 for the 2nd layer, 64 for the 3rd and 4th layers. The classifier consists of 2 FC layers and a softmax layer. After passing the softmax layer, the model outputs the diagnostic label. The detailed model structure is shown in **Figure 6**.

EXPERIMENTS AND RESULTS

We have evaluated the effectiveness of our proposed method in the following experiments. Firstly, we evaluated the quality of the synthetic images generated by 3D reversible GAN. Some image quality metrics were used for the evaluation of the reconstructed images. After that, we compared our classification model with other methods including the VGG and ResNet model in the diagnosis task of AD vs. CN. Finally, we verified the impact of the generated PET data in the classification task of Alzheimer's diagnosis and prediction. Bold values mean the best value in the comparative experiment.

Experiment Setting

Experimental environment: PyTorch framework, Nvidia GTX1080 GPU. During the image reconstruction training, the Adam optimizer (Kingma and Ba, 2014) was used. The reversible blocks were implemented using a modified version of MemCNN (van de Leemput et al., 2018). Peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) (Hore and Ziou, 2010) are used as evaluation metrics of image quality. RMSE, PSNR and SSIM are the most common and widely used objective image evaluation metrics in image reconstruction. RMSE and PSNR measure absolute errors between source and synthetic images, and SSIM measures structural similarities between them. In the classification experiment, we continue to use Adam as the optimizer. The dataset was separated into training part, validation part and testing part. The ratio of training set, validation set, and test set is 7:2:1. ACC, SEN, SPE, AUC are used as evaluation metrics. These four metrics represent accuracy, sensitivity, specificity and area under the curve, respectively. The higher values of these metrics indicate better classification performance. A 10-fold cross-validation approach was employed to measure and to compare the performance of different methods.

Results of Image Reconstruction

Paired ROI Image

From **Figure 7**, we can see that our synthetic PET images are very similar to their corresponding real images. When calculating the

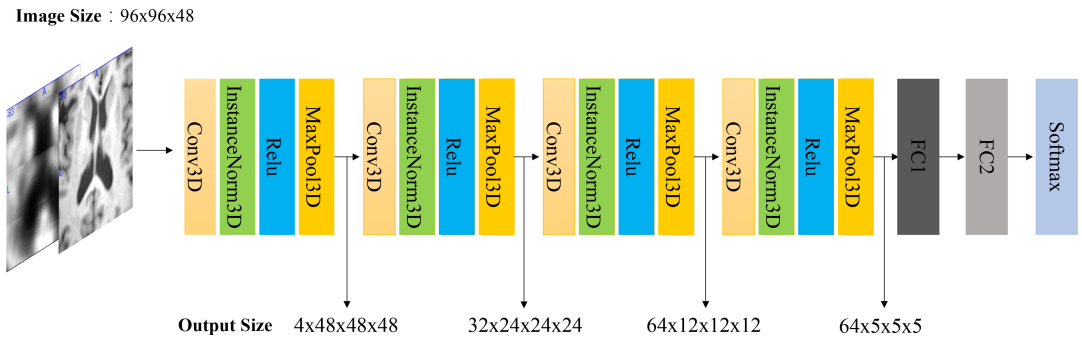


FIGURE 6 | Architecture of our 3D CNN model. Note that if the input is MRI and PET, the input size is 2 × 96 × 96 × 48. The first dimension is the number of channels. FC is a full connected layer. The parameters of FC1 and FC2 are 512 and 2, respectively.

deviation image, we use matplotlib to draw the heat map, and the deviation image is normalized to [0, 255] for display. There are only positive values in the deviation image, and different slices have different ranges, the largest range is [0, 255]. These results indicate that our trained RevGAN model can generate synthetic PET scans with good image quality that are similar to real PET scans. In addition, we have listed the experimental results in some related papers using different methods in recent years for reference. The results are in **Table 2**, which shows that the results of different methods are similar in terms of PSNR in PET image reconstruction. It can be seen from **Table 2** that our method SSIM reached 0.9389 in MRI image reconstruction, achieving the best performance.

Full Image

To compare the performance based on ROI and that based full images, we have performed a further experiment using full images. The registered full image data (221, 257, 221) is scaled using scipy.ndimage.zoom, and the size is resized to (110, 128, 110), and the resolution of the image is reduced. It can be seen from the results of **Figure 7** and **Table 2** that the metric of the generated MR image has decreased significantly. Comparing to the use of ROI, the full MR image contains not only the structural information of the brain tissue but also the structural information such as the skull. Although PET images also have complete head coverage, the signal levels of non-brain tissues are quite different compared with brain tissues, so it creates challenges in alignment between the MR images and the PET image. This will lead to decrease the reconstruction performance of image synthesis.

Results of Disease Diagnosis Prediction

Evaluation of our Proposed 3D CNN Classification Model

We chose 3D-VGG11 and 3D-ResNet10 as the baseline models for comparison. We used the same settings for all models to make fair comparisons. The experimental results are shown in **Tables 3**, **4**. The experiment in **Table 3** used MRI-ROI data, while the experiment in **Table 4** used PET-ROI data. As shown in **Table 3**, we can see that the ACC and SEN of our proposed 3D CNN model achieved 82.00 and 82.29%, respectively. The experimental results are better than the VGG11 and ResNet10 models. The 3D VGG11 model is slightly better than our model in SPE and AUC metrics. As shown in **Table 4**, it can be seen that VGG11 has a very good performance on PET images. The ACC, SEN, and AUC reached 89.11, 90.24, and 92.64%, respectively. It is worth noting that our proposed model is higher than the other two models in the SPE metric. If the age-correction processing of Lin et al. (2018) was performed, the performance of the proposed method was improved. The experimental results show that our four-layer model has better performance than the VGG11 and ResNet10 models which have many layers with much more parameters, indicating the effectiveness of our proposed classification method in the diagnosis of AD.

Results of Diagnosis Using Synthetic Data

The model used in all multi-modal experiment is our 3D CNN model. We simulated the absence of data and tested the impact of the generated PET on the diagnosis and prediction of AD (Tong et al., 2017) used non-linear graph fusion to fuse the features of different modalities. To take advantage of information about the disease, we adopted a framework to integrate multi-modality information based on our proposed 3D CNN model. In this part of the experiment, MR and PET images are used as two parallel channels. After then, the paired MR image and PET image are stacked as a 4D image. We also employed the above four metrics for performance evaluation of AD diagnosis. Disease classification results were reported in **Tables 5**–**8**.

As shown in **Table 5**, the method used real data got the best metrics in terms of ACC, SPE, and AUC. Note that "MRI + 100% synthetic PET" has the highest value in SEN. A similar situation can be seen in **Tables 6**–**8**. The method used real data got the best metrics in terms of ACC, SEN, while using 100% synthetic PET has achieved the highest AUC in **Table 6**. In **Table 7**, the method using real data obtained the best indicators in terms of ACC, SPE, and AUC. Please note that "PET + 100% synthetic MRI" has the highest SEN value. In **Table 8**, the method used real data got the best metrics in terms of ACC, SEN, while using 100% synthetic MRI has achieved the highest SPE and AUC. As shown in **Tables 5**–**8**, after using synthetic data, there is a higher classification score. This shows that our method is effective. The experimental ROC curves are shown in **Figures 8**, **9**. Moreover, the relevant experimental results using full image are shown in

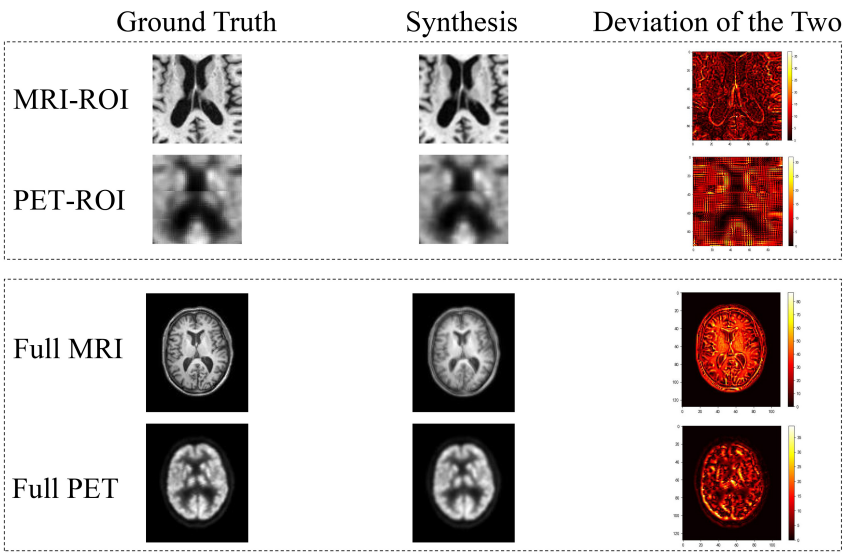


FIGURE 7 | Deviation between real image and synthetic images. In the deviation image, the yellower the color, the greater the error, and the red and black areas represent smaller deviation.

TABLE 2 | Comparison with others.

		Synthetic PET		Synthetic MRI	
		PSNR	**SSIM**	**PSNR**	**SSIM**
Pan et al., 2020	HGAN	**30.24**	0.6945	26.07	0.6683
Pan et al., 2019	FGAN	29.62	0.6817	25.10	0.6404
Hu et al., 2019	Adversarial U-Net (2D slice)	25.13	–	–	–
Pan et al., 2018	3D-cGAN	24.49	–	–	–
The proposed method (ROI)	RevGAN	29.34	0.8034	**29.81**	**0.9389**
The proposed method (Full Data)	RevGAN	29.42	**0.8176**	24.97	0.6746

The quality of the reconstructed image. Results of image synthesis achieved by different methods. Just as a reference, because the selected data and data preprocessing are different. Bold values mean the best value in the comparative experiment.

Tables 9, **10**, and the ROC curve is shown in **Figure 10**. In this study, for the classification tasks using full image, resampled images at a lower resolution were used instead of using multiple patches as (Pan et al., 2018). We think that not all patches within the whole image were affected by the disease changes of AD. Some patches from AD subjected may be not affected by AD and their diagnostic label may be fuzzy. Therefore, if the selected patches are not accurate, it will lead to poor performance. In addition, compared with using a full image, using multiple patches will lose the spatial position information between different brain regions.

DISCUSSION

In the classification task, we propose a 3D CNN model with only four convolutional layers to achieve T1-MRI and FDG-PET AD diagnosis. The hippocampal area was used as ROI for input as it is the most frequently studied and is thought to be of the highest related region of AD. Through our experiments, we found that the proposed four-layer network has a lightweight structure but with competitive performance in AD classification tasks. Especially, we use small-sized kernels in the first convolutional layer to prevent early spatial downsampling in contrast to most standard architectures for image classification. For example, ResNet uses relatively large kernel sizes and strides in the first layer, which dramatically reduces the spatial dimension of their inputs. This can accelerate the computation. However, the loss of a large amount of information at the beginning layer may adversely affect the results, especially in medical images (Liu et al., 2020). In addition, the input size is only $96 \times 96 \times 48$ due to the use of ROI-based data in our experiments. Therefore, premature downsampling can cause severe performance degradation. It can be seen from **Table 5** that the performance of "50% synthetic + 50% real" is worse than "100% real" and "100 synthetic." This indicates that the complexity of the data will affect the performance, because the distribution of synthetic data and actual data may be slightly different, and mixing the two types will make it difficult to train the model. It can be seen from **Table 4** that very good performance can be achieved using only PET data. From the **Table 6**, it is strange that best SPE was achieved with "MRI+50%PET+50%synthetic PET," and best AUC is achieved with imputed PET (better than when full real data

TABLE 3 | Results (%) of the models trained from only MRI data for CN vs. AD task.

Model	ACC	SEN	SPE	AUC
3D-VGG11	81.19	79.27	**83.45**	**83.67**
3D-ResNet10	80.87	79.63	82.23	81.21
3D CNN (4 layers)	**82.00**	**82.29**	81.69	81.76

Bold values mean the best value in the comparative experiment.

TABLE 4 | Results (%) of the models trained from only PET data for CN vs. AD task.

Model	ACC	SEN	SPE	AUC
3D-VGG11 (Huang et al., 2019)	**89.11**	**90.24**	87.77	**92.69**
3D-ResNet10	86.26	86.56	85.94	84.48
3D CNN (4 layers)	88.77	89.11	**88.41**	87.11

Bold values mean the best value in the comparative experiment.

TABLE 5 | Diagnosis results (%).

	ACC	SEN	SPE	AUC
MRI+PET (Real data)	**89.26**	82.69	**96.48**	**90.98**
MRI+PET (Miss 50% data)	84.64	85.82	83.45	83.92
MRI+50%PET+50% synthetic PET	87.63	87.07	88.24	87.07
MRI+ the other 50% synthetic PET	87.99	87.76	88.24	87.36
MRI+100%synthetic PET	89.05	**90.48**	87.50	87.92

In this AD vs. CN classification experiment, all MR images are real, but PET images were divided into five cases. The data used in the experiment is an ROI image. "MRI+ the other 50% synthetic PET" is to replace the original real data in "MRI+50%PET+50% synthetic PET" with synthetic data and change the synthetic data to real data. Bold values mean the best value in the comparative experiment.

TABLE 6 | Diagnosis results (%).

	ACC	SEN	SPE	AUC
MRI+PET (Real data)	**72.84**	**88.89**	52.78	68.30
PET+MRI (Miss 50% data)	66.87	67.05	66.67	68.95
MRI+50%PET+50% synthetic PET	69.28	55.68	**84.62**	70.29
MRI+ the other 50% synthetic PET	68.49	84.62	50.00	67.53
MRI+100%synthetic PET	71.23	74.36	67.65	**73.66**

In this pMCI vs. sMCI classification experiment, all MR images are real, but PET images were divided into five cases. The data used in the experiment is an ROI image. Bold values mean the best value in the comparative experiment.

TABLE 7 | In this AD vs. CN classification experiment, all PET images are real, but MR images were divided into five cases.

	ACC	SEN	SPE	AUC
MRI+PET (Real data)	**89.26**	82.69	**96.48**	**90.98**
MRI+PET (Miss 50% data)	84.64	85.82	83.45	83.92
PET+50%MRI+50% synthetic MRI	87.12	81.68	92.48	85.59
PET+ the other 50% synthetic MRI	86.71	87.73	85.51	84.92
PET+100%synthetic MRI	88.64	**91.60**	85.71	87.36

The data used in the experiment is an ROI image. Bold values mean the best value in the comparative experiment.

TABLE 8 | In this sMCI vs. pMCI classification experiment, all PET images are real, but MR images were divided into five cases.

	ACC	SEN	SPE	AUC
MRI+PET (Real data)	**72.84**	**88.89**	52.78	68.30
PET+MRI (Miss 50% data)	66.87	67.05	66.67	68.95
PET+50%MRI+50% synthetic MRI	67.81	83.33	50.00	67.87
PET+ the other 50% synthetic MRI	68.67	67.05	70.51	69.81
PET+100% synthetic MRI	71.18	69.07	**73.97**	**70.80**

The data used in the experiment is an ROI image. Bold values mean the best value in the comparative experiment.

are available). In lots of previous studies, the use of single-modal PET has been able to achieve very good results in the classification of AD. In this work, all multi-modal experiments use the four-layer 3D CNN model we designed, and the best result of single-modal PET is to use the 3D VGG11 of Huang et al. (2019). The models used are inconsistent (**Tables 4**, **5**). These four indicators describe the performance of the model from different aspects, rather than absolute positive correlation. Therefore, due to different experimental settings, some evaluation metrics will fluctuate. In the comparative experiments using real data and synthetic data, the number of data samples is small. AD, NC, pMCI, and sMCI are 647, 707, 326, and 396, respectively. The numbers of samples in different categories are imbalanced. This makes the model biased toward a certain type of sample in order to reduce its own loss, which causes fluctuations in some metrics. When data is missing or synthetic data is used, this situation is more likely to happen.

From **Tables 9**, **10**, it can be seen that although the image resolution becomes lower when using the full image, it still achieves a good classification score. The analysis in "Full Image" shows that the structural and functional information of brain tissue can be mapped very well, but it is difficult to map the structural information such as the skull of the MR image from the PET image. Therefore, it does not perform well when synthesizing a full MRI image from a PET image (**Table 2**). This problem can be avoided if only the hippocampus ROI is used. Although the quality of the synthesized MR image is not high, the classification result of using the synthesized MR image is not bad (**Tables 9**, **10**). This shows that the model mainly focuses on changes in brain tissue structure (such as hippocampus atrophy, etc.) when diagnosing and predicting diseases, and the structure of irrelevant areas such as the skull has little effect on the results.

In the image synthesis task, reversible blocks are very memory-efficient. Therefore, by adopting the generator together with the stability of reversible architecture, this allows us to train deeper networks using only modest computational resources. Increasing the depth of the reversible block improves the non-linear fitting ability of the model, which can generate higher quality images. Moreover, reversible architecture has been demonstrated to yield superior results when trained using fewer training data in the work of Chang et al. (2018). It can be seen from **Table 2** that compared with our work, Pan et al. (2018, 2019, 2020) performed more preprocessing steps to obtain better alignment

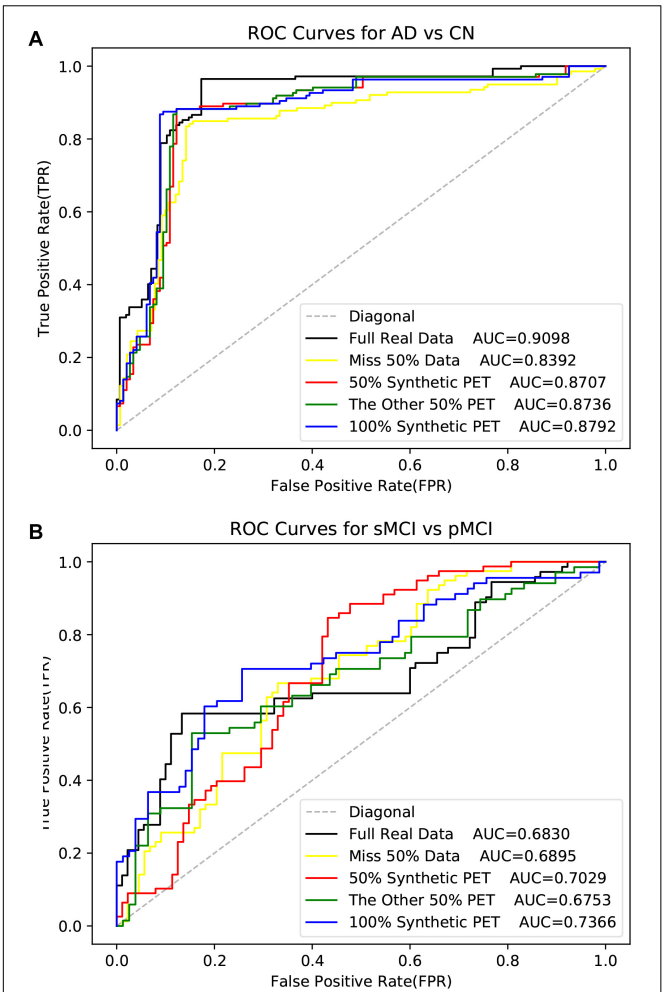


FIGURE 8 | The ROI data is used here. The ROC curves in the experiment of disease classification with the synthetic PET image.

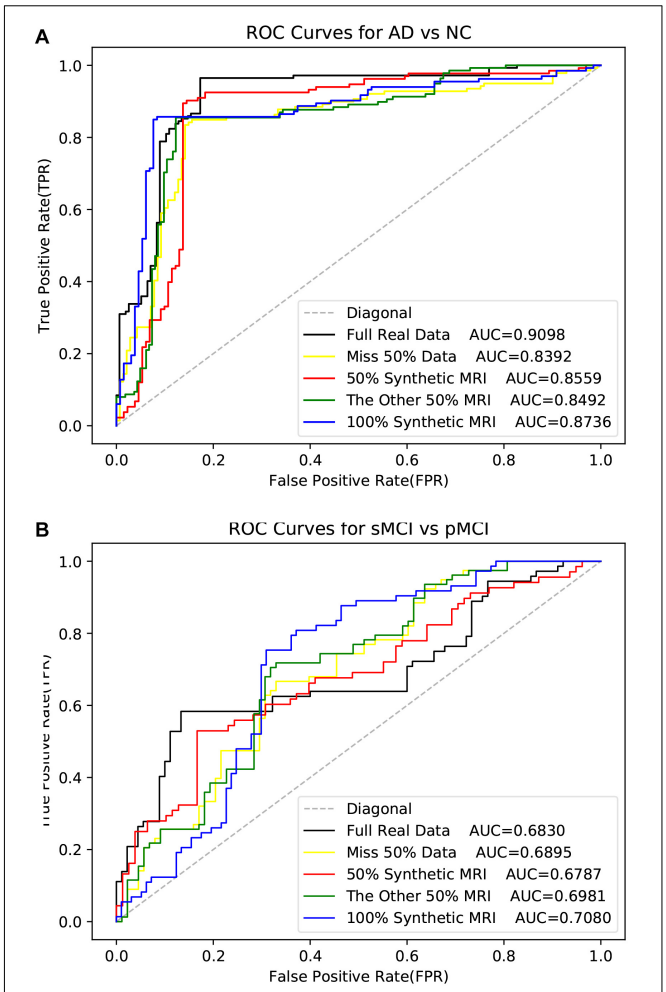


FIGURE 9 | The ROC curves in the experiment of disease classification with the synthetic MR image. The ROI data is used here.

between MR image and PET images, thus resulting in a good reconstruction results. Despite their advantage on the alignment, our proposed method are still superior to Pan et al. (2018) work on the reconstruction quality of synthetic images in some metrics as shown in **Table 2**. It is worth noting that too much preprocessing may lose more original information and affect the reliability of the experiment. When registering the PET images to the corresponding MRI, we use nearest neighbor interpolation (Olivier and Hanqiang, 2012). This leads to moiré pattern (Oster and Nishijima, 1963) on the PET image. Compared with the real image, the synthesized PET image has reduced moiré pattern (see **Figure 7**). Related experimental results are shown in **Table 2**. From **Tables 5–8**, we can see that the performance of the method using our synthetic data is superior to those using missing data. The experimental results show that our synthetic PET is beneficial for disease diagnosis. If our synthetic data does not contain disease information, it may hamper the performance of AD diagnosis, leading to worse performance.

TABLE 9 | The AD vs. CN classification experiment uses synthetic full data and real full data.

	ACC	SEN	SPE	AUC
MRI+PET	**92.28**	90.38	**94.37**	**92.76**
MRI+100% synthetic PET	91.95	89.74	94.37	92.51
PET+100% synthetic MRI	90.77	**90.58**	90.98	91.60

Bold values mean the best value in the comparative experiment.

The main limitations of this study and future work are as follows. In this study, since the hippocampus is an appropriate ROI selection for the AD diagnosis, we can use the ROI block that removes the redundant information in the data to diagnose and predict the disease and improve performance. When applied to other diseases without a specific target ROI, it is difficult to take advantage of the ROI-based method. Moreover, it would be interesting to show the biologically meaningful measurements of the synthetic data over the real data. In this work, the synthetic PET data were used for AD diagnosis and focused more on the disease changes of AD, rather than the real FDG

TABLE 10 | The sMCI vs. pMCI classification experiment uses synthetic full data and real full data.

	ACC	SEN	SPE	AUC
MRI+PET	**74.10**	**75.00**	73.08	**76.60**
MRI+100% synthetic PET	73.78	64.52	**85.92**	75.09
PET+100% synthetic MRI	73.49	73.86	73.08	76.03

Bold values mean the best value in the comparative experiment.

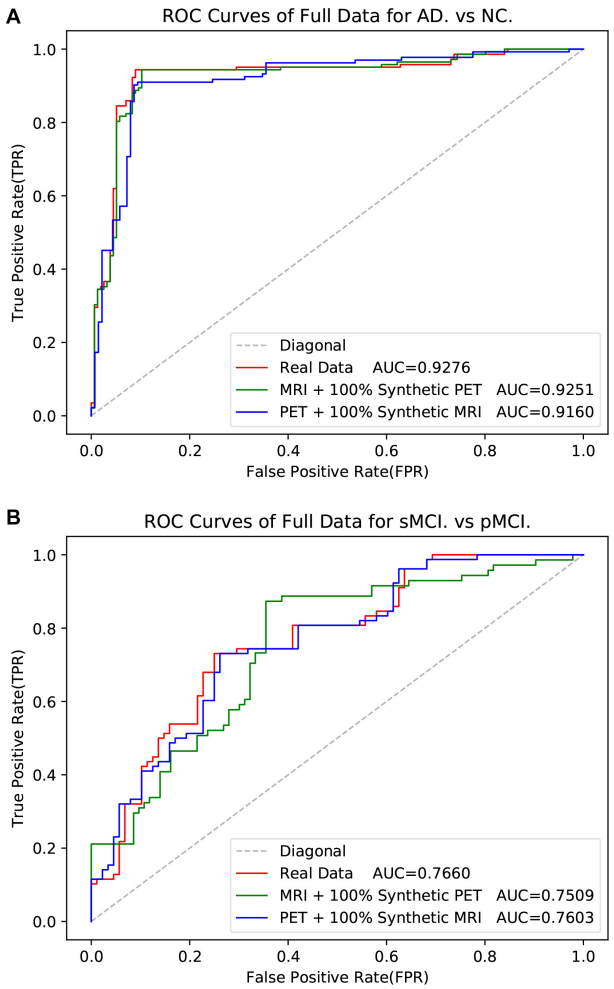


FIGURE 10 | The ROC curves in the experiment of classification with the full data.

metabolic changes. The aim of the reconstruction is to improve the classification of AD in the scenario of missing data. The detailed investigation of reconstructing biologically meaningful synthetic data will be carried out in future. Last but not the least, in **Tables 5**, **6**, the ratio was set to 50% (half synthetic PET data) and 100% (all synthetic PET data) in our experiments. By adding more synthetic PET data, the accuracy has been improved by comparing the new results in **Tables 5**, **6**. The effect of the ratio parameter in the random selection of the synthetic PET data will be carefully studied in future work. Finally, the synthesized PET images may be helpful for improving the AD diagnosis performance, while they do not reflect the real metabolic changes of the FDG PET imaging. In **Table 3**, we have compared the performance of three different CNN models on the AD vs. CN task based on MRI data. Comparing **Tables 3**, **5**, we can find that the performance of MRI+synthetic PET and MRI+PET (Real data) is superior than that of MRI alone due to the complementary information. But, we also believe that there is not a strong relationship between MR and PET in all scenarios, and a good mapping between MRI and PET needs to meet some conditions. If the synthetic model that we trained in the AD diagnosis scene is directly applied to other diseases, it may not work. For example, if we want to apply our method in the scene of tumor diagnosis, a large number of paired MR and PET must be used to retrain the model so that the model can find the relationship between MR and PET in the scene of tumor diagnosis. In this study, hippocampal atrophy in AD patients will be reflected in the structure and functional metabolism of the hippocampus. However, there are many other diseases that may not affect tissue structure and tissue function metabolism at the same time, so it is difficult to find the relationship between MRI and PET. To sum up, whether the disease can cause changes in tissue structure and metabolic function simultaneously may be a condition for MR and PET to be able to map well. Therefore, further research needs to be explored.

CONCLUSION

To conclude, we proposed a 3D reversible GAN for imputing those missing data to address the issue of missing data. Specifically, we have also presented a novel 3D CNN architecture to perform classification for AD diagnosis. Moreover, we tested the impact of the synthetic data in the classification task of AD by simulating missing data. During the experiment to evaluate the impact of synthetic data, the multi-modal fusion method by channel fusion (MR images and PET images were stacked into 4D images) is selected. Experiments on the ADNI dataset demonstrate that our method generates reasonable neuroimages. Through the experimental results, we can also find the following three conclusions: First, we can find that the structural and functional information of brain tissue can be mapped well, but it is difficult to map the structure information such as the skull of the MR image from the PET image. Second, the classification model is mainly based on the brain tissue area in the neuroimaging and is not sensitive to the skull and other structures. Third, when the data is missing, the performance of AD diagnosis and MCI conversion prediction can be significantly improved using our method.

THE ALZHEIMER's DISEASE NEUROIMAGING INITIATIVE

The Alzheimer's Disease Neuroimaging Initiative is funded by the National Institute on Aging, the National Institute of

Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

As per ADNI protocols, all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The ADNI data collection was carried out after obtaining written informed consent from the participants. More details can be found at adni.loni.usc.edu.

AUTHOR CONTRIBUTIONS

TT and MD supervised and managed the research, and revised the manuscript. WaL leads the implementations, experiments, and wrote the manuscript. WeL co-lead the work and co-investigated the work. GC, HZ, and QG provided support to the work of experiments and revised the manuscript. YH gave support in data preprocessing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China under Grant 61901120 and Grant 61802065 and in part by the Science and Technology Program of Fujian Province of China under Grant 2019YZ016006, in part by the Foundation of Educational and Scientific Research Projects for Young and Middle-aged Teachers of Fujian Province under Grant JAT200471, and in part by the High-level Talent Project of Xiamen University of Technology under Grant YKJ20013R.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

REFERENCES

BenTaieb, A., and Hamarneh, G. (2017). Adversarial stain transfer for histopathology image analysis. *IEEE Trans. Med. Imaging* 37, 792–802. doi: 10.1109/tmi.2017.2781228

Bi, L., Kim, J., Kumar, A., Feng, D., and Fulham, M. (2017). "Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs)," in *Proceedings of the Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, eds M. J. Cardoso, et al. (Cham: Springer), 43–51. doi: 10.1007/978-3-319-67564-0_5

Calhoun, V. D., and Sui, J. (2016). Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1, 230–244. doi: 10.1016/j.bpsc.2015.12.005

Chang, B., Meng, L., Haber, E., Ruthotto, L., Begert, D., and Holtham, E. (2018). "Reversible architectures for arbitrarily deep residual neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, LA.

Dickerson, B. C., Goncharova, I., Sullivan, M., Forchetti, C., Wilson, R., Bennett, D., et al. (2001). MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiol. Aging* 22, 747–754. doi: 10.1016/s0197-4580(01)00271-8

Dinh, L., Krueger, D., and Bengio, Y. (2014). Nice: non-linear independent components estimation. arXiv [Preprint]. Available online at: https://arxiv.org/abs/1410.8516

Dong, C., Loy, C. C., He, K., and Tang, X. (2014). "Learning a deep convolutional network for image super-resolution," in *Proceedings of the European Conference on Computer Vision*, eds D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer), 184–199.

Fan, Y., Gur, R. E., Gur, R. C., Wu, X., Shen, D., Calkins, M. E., et al. (2008). Unaffected family members and schizophrenia patients share brain structure patterns: a high-dimensional pattern classification study. *Biol. Psychiatry* 63, 118–124. doi: 10.1016/j.biopsych.2007.03.015

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). "Synthetic data augmentation using GAN for improved liver lesion classification," in *Proceedings of the IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, Washington, DC, 289–293.

Gomez, A. N., Ren, M., Urtasun, R., and Grosse, R. B. (2017). "The reversible residual network: backpropagation without storing activations," in *Proceedings of the Advances in Neural Information Processing Systems*, Long Beach, CA, 2214–2224.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal, 2672–2680.

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*, Las Vegas, NV, 770–778.

Heckemann, R. A., Ledig, C., Gray, K. R., Aljabar, P., Rueckert, D., Hajnal, J. V., et al. (2015). Brain extraction using label propagation and group agreement: Pincram. *PLoS One* 10:e0129211. doi: 10.1371/journal.pone.0129211

Hore, A., and Ziou, D. (2010). "Image quality metrics: PSNR vs. SSIM," in *Proceedings of the 20th International Conference on Pattern Recognition*, Istanbul, 2366–2369.

Hu, S., Yuan, J., and Wang, S. (2019). "Cross-modality synthesis from MRI to PET using adversarial U-net with different normalization," in *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, Shenzhen, 1–5. doi: 10.1109/ICMIPE47306.2019.9098219

Huang, Y., Xu, J., Zhou, Y., Tong, T., Zhuang, X., and The Alzheimer's Disease Neuroimaging Initiative [ADNI] (2019). Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. *Front. Neurosci.* 13:509. doi: 10.3389/fnins.2019.00509

Jack, C. R. Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691.

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. arXiv [Preprint]. Available online at: https://arxiv.org/pdf/1412.6980.pdf

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.

Ledig, C., Heckemann, R. A., Hammers, A., Lopez, J. C., Newcombe, V. F., Makropoulos, A., et al. (2015). Robust whole-brain segmentation: application to traumatic brain injury. *Med. Image Anal.* 21, 40–58. doi: 10.1016/j.media.2014.12.003

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., et al. (2014). "Deep learning based imaging data completion for improved brain disease diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer), 305–312. doi: https://doi.org/10.1007/978-3-319-10443-0_39

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., et al. (2018). Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front. Neurosci.* 12:777. doi: 10.3389/fnins.2018.00777

Liu, M., Gao, Y., Yap, P.-T., and Shen, D. (2017). Multi-hypergraph learning for incomplete multimodality data. *IEEE J. Biomed. Health Inform.* 22, 1197–1208. doi: 10.1109/jbhi.2017.2732287

Liu, S., Yadav, C., Fernandez-Granda, C., and Razavian, N. (2020). "On the design of convolutional neural networks for automatic detection of Alzheimer's disease", in *Proceedings of the Machine Learning for Health Workshop, (PMLR)*, Durham, NC, 116, 184–201.

Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., et al. (2017). "Medical image synthesis with context-aware generative adversarial networks," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, eds L. Maier-Hein, A. Franz, P. Jannin, S. Duchesne, M. Descoteaux, and D. L. Collins (Cham: Springer), 417–425. doi: 10.1007/978-3-319-66179-7_48

Olivier, R., and Hanqiang, C. (2012). Nearest neighbor value interpolation. *Int. J. Adv. Comput. Sci. Appl.* 3, 25–30. doi: 10.14569/IJACSA.2012.030405

Oster, G., and Nishijima, Y. (1963). Moiré patterns. *Sci. Am.* 208, 54–63.

Pan, Y., Liu, M., Lian, C., Xia, Y., and Shen, D. (2019). "Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, eds D. Shen, et al. (Cham: Springer), 137–145. doi: 10.1007/978-3-030-32248-9_16

Pan, Y., Liu, M., Lian, C., Xia, Y., and Shen, D. (2020). Spatially-constrained Fisher representation for brain disease identification with incomplete multi-modal neuroimages. *IEEE Trans. Med. Imaging* 39, 2965–2975. doi: 10.1109/tmi.2020.2983085

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., and Shen, D. (2018). "Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, eds A. F. Frangi, C. Davatzikos, G. Fichtinger, C. Alberola-López, and J. A. Schnabel (Cham: Springer), 455–463. doi: 10.1007/978-3-030-00931-1_52

Salvatore, C., Cerasa, A., Battista, P., Gilardi, M. C., Quattrone, A., and Castiglioni, I. (2015). Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Front. Neurosci.* 9:307. doi: 10.3389/fnins.2015.00307

Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L., Trojanowski, J., et al. (2009). MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain* 132, 1067–1077. doi: 10.1093/brain/awp007

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv [Preprint]. Available online at: https://dblp.org/rec/journals/corr/SimonyanZ14a.html

Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D., and The Alzheimer's Disease Neuroimaging Initiative (2017). Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recognit.* 63, 171–181. doi: 10.1016/j.patcog.2016.10.009

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2016). Instance normalization: the missing ingredient for fast stylization. *arXiv* [Preprint]. Available online at: https://arxiv.org/pdf/1607.08022.pdf

van de Leemput, S. C., Teuwen, J., and Manniesing, R. (2018). "Memcnn: a framework for developing memory efficient deep invertible networks," in *Proceedings of the 6th International Conference on Learning Representations, Workshop, (ICLR)*, Vancouver.

van der Ouderaa, T. F., and Worrall, D. E. (2019). "Reversible gans for memory-efficient image-to-image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 4720–4728.

Wee, C.-Y., Yap, P.-T., Zhang, D., Wang, L., and Shen, D. (2014). Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. *Brain Struct. Funct.* 219, 641–656. doi: 10.1007/s00429-013-0524-8

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., et al. (2019). Predicting PET-derived demyelination from multimodal MRI using sketcher-refiner adversarial training for multiple sclerosis. *Med. Image Anal.* 58:101546. doi: 10.1016/j.media.2019.101546

Yu, Y., Gong, Z., Zhong, P., and Shan, J. (2017). "Unsupervised representation learning with deep convolutional neural network for remote sensing images," in *Proceedings of the Springer International Conference on Image and Graphics*, Shanghai, 97–108. doi: 10.1007/978-3-319-71589-6_9

Zhao, M., Wang, L., Chen, J., Nie, D., Cong, Y., Ahmad, S., et al. (2018). "Craniomaxillofacial bony structures segmentation from MRI with deep-supervision adversarial learning," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, eds A. F. Frangi, G. Fichtinger, J. A. Schnabel, C. Alberola-López, and C. Davatzikos (Cham: Springer), 720–727. doi: 10.1007/978-3-030-00937-3_82

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 2223–2232.

Conflict of Interest: QG and TT was partially employed by the company Imperial Vision Technology.

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.

Copyright © 2021 Lin, Lin, Chen, Zhang, Gao, Huang, Tong, Du and the Alzheimer's Disease Neuroimaging Initiative. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 20 April 2021
doi: 10.3389/fnins.2021.630747

Voxel-Wise Feature Selection Method for CNN Binary Classification of Neuroimaging Data

*Domenico Messina[1], Pasquale Borrelli[1], Paolo Russo[2], Marco Salvatore[1] and Marco Aiello[1]**

[1] IRCCS SDN, Naples, Italy, [2] Dipartimento di Fisica "Ettore Pancini", Università Degli Studi di Napoli "Federico II" – Complesso Universitario di Monte Sant'Angelo, Naples, Italy

Voxel-wise group analysis is presented as a novel feature selection (FS) technique for a deep learning (DL) approach to brain imaging data classification. The method, based on a voxel-wise two-sample *t*-test and denoted as *t*-masking, is integrated into the learning procedure as a data-driven FS strategy. t-Masking has been introduced in a convolutional neural network (CNN) for the test bench of binary classification of very-mild Alzheimer's disease vs. normal control, using a structural magnetic resonance imaging dataset of 180 subjects. To better characterize the *t*-masking impact on CNN classification performance, six different experimental configurations were designed. Moreover, the performances of the presented FS method were compared to those of similar machine learning (ML) models that relied on different FS approaches. Overall, our results show an enhancement of about 6% in performance when *t*-masking was applied. Moreover, the reported performance enhancement was higher with respect to similar FS-based ML models. In addition, evaluation of the impact of *t*-masking on various selection rates has been provided, serving as a useful characterization for future insights. The proposed approach is also highly generalizable to other DL architectures, neuroimaging modalities, and brain pathologies.

Keywords: deep learning, feature selection, neuroimaging, statistical parametric mapping, t-masking, Alzheimer's disease, magnetic resonance imaging, brain disorders

OPEN ACCESS

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Muhammad Fazal Ijaz,
Sejong University, South Korea
Siyuan Lu,
University of Leicester,
United Kingdom

****Correspondence:***
Marco Aiello
marco.aiello@synlab.it

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *18 November 2020*
Accepted: *26 February 2021*
Published: *20 April 2021*

Citation:
Messina D, Borrelli P, Russo P, Salvatore M and Aiello M (2021) Voxel-Wise Feature Selection Method for CNN Binary Classification of Neuroimaging Data. Front. Neurosci. 15:630747. doi: 10.3389/fnins.2021.630747

INTRODUCTION

During the last decade, technological advancements and the availability of large amounts of labeled data (Aiello et al., 2019) fostered neuroimaging research's development (Traverso et al., 2020). In this context, machine learning (ML) algorithms played a relevant role (Lundervold and Lundervold, 2019; Chatterjee et al., 2020; Salmanpour, 2020; Traverso et al., 2020). Indeed, ML approaches aimed to automatically recognize meaningful patterns undetectable with human perception. ML methods enabled the potential development of computer-aided diagnosis and decision support systems for diagnosis and clinical management of a high number of diseases (Yassin et al., 2018). Moreover, ML algorithms were successfully applied to perform different tasks, like image classification, object detection, and image segmentation (Dora et al., 2017).

Concerning classification tasks, some ML studies focused on support vector machines (SVMs) and random forests for binary classification of pathological versus healthy conditions

(Davatzikos, 2019) by using features manually extracted from raw data or features learned, in turn, by other simple ML models (Lundervold and Lundervold, 2019; Singh et al., 2020).

In the context of ML algorithms, deep learning (DL) gained considerable attention among the scientific community for medical imaging applications, hence also in the neuroimaging field (Ding et al., 2019; Jo et al., 2019; Spasov et al., 2019; Zhu et al., 2019). Unlike the classical ML approaches, where prior knowledge of the domain was fundamental, DL utilizes deep neural networks to automatically discover and extract useful features directly from the data (LeCun et al., 2015). Deep neural networks, in particular convolutional neural networks (CNNs), outperformed previous state-of-the-art ML approaches (Litjens et al., 2017; Shen et al., 2017; Brinker et al., 2019). In the context of neuroimaging, DL was applied to classify psychiatric and neurological disorders, which tend to be associated with subtle and diffuse neuroanatomical and neurofunctional abnormalities (Vieira et al., 2017).

Typically, ML approaches widely used feature selection (FS) techniques, i.e., the process of choosing a subset of relevant features for use in model design. The FS goal is finding the best feature subset that yields the minimum generalization error, enhancing the model's performance (Vergara and Estevez, 2014). FS methods improve the generalization by avoiding the curse of dimensionality and by reducing overfitting (Singh et al., 2016). Furthermore, FS techniques are increasingly employed since they allow a simplification of learning models, facilitating human interpretability. They also have the advantage of diminishing training times (Venkatesh and Anuradha, 2019). However, these ML approaches share the requirement of high domain technical knowledge, thus limiting their applicability to specific classification tasks (LeCun et al., 2015). Data-driven approaches provided fundamental improvements for the generalization of ML algorithms to deal with this issue. For example, Ijaz et al. (2020) utilized a data-driven prediction model with outlier detection based on random forests and used a chi-squared FS method. Instead, Ijaz et al. (2018) uses an *Information Gain* technique to evaluate the features' significance.

In general, DL algorithms did not require FS approaches since DL automatically discovers the intricate structure of imaging features in large datasets during the training step (LeCun et al., 2015). However, FS techniques could improve the DL performance for a classification task, mostly in the case of limited data availability (Chen et al., 2020; Raj et al., 2020). Indeed, in the domain of computational neuroscience, the available image datasets are often characterized by the size of features much larger than the number of examples (the curse of dimensionality) (Abraham et al., 2014; Mwangi et al., 2014). In the context of high-dimensional classification, in general, FS approaches have been proven suitable in several biomedical applications, also in DL employments (Fan and Fan, 2008; Suk et al., 2016; Raj et al., 2020). However, in deep neural network models, FS methods are still poorly investigated (Raj et al., 2020).

Voxel-wise two-sample *t*-tests were demonstrated to be effective in revealing brain areas associated with statistically significant differences between groups related to a multitude of neurological pathologies, for example, Alzheimer's disease (AD) (Senjem et al., 2005), multiple sclerosis (Audoin et al., 2010), amyotrophic lateral sclerosis (Sage et al., 2007), childhood absence epilepsy (Pardoe et al., 2008), and schizophrenia (Asami et al., 2014).

To date, limited attention has been paid to deriving statistical significance maps of imaging patterns to establish if a voxel or a deep feature constitutes a significant contributor to a DL model (Davatzikos, 2019). In general, FS approaches based on statistical saliency are introduced in association with simpler classifiers than those in DL models. For example, in Tohka et al. (2016), several data-driven FS and classification methods are proposed for the whole-brain voxel-based classification of AD vs. normal control (NC) subjects.

The inclusion of an FS criterion into a DL pipeline represents an attractive and non-trivial challenge due to the non-linear nature of DL models. Moreover, this criterion should be automatic, generalizable, and data driven, to maintain the critical aspect of DL: to have multiple layers of features not designed by human engineers, but learned from data using a general-purpose learning procedure (LeCun et al., 2015).

In this work, we introduced a novel FS technique for CNN architectures. Our primary aim was to verify the effective improvement of a binary classification task by including a data-driven FS approach based on a voxel-wise test of statistical significance. The choice of applying FS to a CNN algorithm was motivated by the reduction of the sample dimensionality without losing relevant information. Indeed, the image dimensionality (e.g., the number of voxels) represents a bottleneck that affects the CNN routines' training for 3D medical imaging classification procedures (Vieira et al., 2017). We aimed to demonstrate the *t*-masking approach's feasibility by assessing its impact on a 3D CNN model's performance. For this purpose, we choose, as a test bench, the classification task of AD vs. NC subjects, widely discussed in the literature. The main reasons for this choice were as follows: (i) It is demonstrated that CNNs have a good baseline level of accuracy performance for datasets of a few hundreds of samples, also in the early stage of pathology (Vieira et al., 2017) and (ii) it is demonstrated that there are subtle region-specific anatomical alterations in AD brains, viewable via structural T1-weighted (T1-w) magnetic resonance imaging (MRI) (Fox and Freeborough, 1997; Ling et al., 2013) and, likely, detectable by a voxel-wise *t*-test. In particular, the classification task consisted of automatically recognizing subjects with very mild AD vs. NC, a task more challenging compared to the AD-vs.-NC classification. However, the task under consideration has a higher clinical interest for early diagnosis development.

In the AD-vs.-NC classification task context, a massive quantity of studies based on ML and DL methods was published. Early studies on AD classification proposed a classification pipeline that started with an FS and extraction steps to obtain useful information to feed on a multivariate pattern classification algorithm (Mateos-Pérez et al., 2018). One of the most employed classifiers was SVM, requiring kernels that transform input data and act as a similarity measure for the classification task. Referring to the FS strategy, Chaves et al. (2009) used an FS based on a *t*-test to select regions of interest, reducing the dimensionality of input data; Asim et al. (2018) utilized a brain

atlas; Moradi et al. (2015) used a simpler ML algorithm as regularized logistic regression; Suk et al. (2016) proposed a weighted sparse multi-task learning method, and Zhang et al. (2014) adopted a principal component analysis (PCA) method. Fan et al. (2007) performed a combination of group analysis on adaptive regional elements with an SVM classifier. Ensemble methods (like the random forest) are also employed for the multimodal classification of AD (Gray et al., 2013).

DL methods have been applied to AD-vs.-NC classification tasks, achieving the highest classification performance, especially when using multimodality data or when combined with other learning approaches (Jo et al., 2019). Spasov et al. (2019) developed a parameter-efficient 3D CNN with a dual-learning approach to predict the conversion from mild cognitive impairment (MCI) to AD. Khagi et al. (2019) used transfer learning to enhance the performance of their CNN. Afzal et al. (2019) proposed an augmentation technique to balance the dataset and improve classification performance, increasing the training set's sample size. Suk et al. (2016) presented a DL-based latent feature representation with a stacked autoencoder, while Choi et al. (2018) introduced PET images as input in a 3D CNN to predict conversion from MCI to AD.

Considering the analyzed literature, to the best of our knowledge, this is the first study in the neuroimaging field developing a DL model with an FS based on voxel-wise statistical analysis for image binary classification.

MATERIALS AND METHODS

Dataset

Data from the OASIS-3 release, which consisted of a longitudinal neuroimaging, clinical, and cognitive study of normal aging and AD (LaMontagne et al., 2019), were used. In particular, we selected T1-w MRI data acquired with a 3-T magnetic resonance (MR) scanner. LaMontagne et al. (2019) reported detailed information on recruitment criteria, imaging acquisition protocols, used scanners, and clinical/neurological assessment. We used the Clinical Dementia Rating (CDR), as provided by OASIS-3, to select AD subjects with very mild AD (i.e., CDR = 0.5). We selected structural T1-w MR images from 90 AD subjects with an MR session made at the time close to AD diagnosis with CDR = 0.5. To obtain a balanced dataset of both AD and NC groups, 90 NC participants, matched for age and sex with the selected AD group, were randomly included by selecting the T1-w images related to the first available sessions, resulting in a dataset with 180 samples.

Preprocessing

The preprocessing procedures consisted of spatial and intensity normalization of neuroimaging data.

Spatial normalization was performed by registering the T1-w images to the Montreal National Institute (MNI)-152 (1 mm) standard space template using the FSL routines (FMRIB-FSL package v. 6.0.0). In particular, we used the FNIRT routine to non-linearly register the T1-w images to the MNI-152 template, including a preliminary linear registration step (FLIRT of FMRIB-FSL) of the brain-extracted T1-w images. The skull stripped T1-w images were generated with a 3D CNN approach, using the DeepBrain Extractor, a Python tool that runs a pre-trained 3D U-net available at https://github/iitzco/deepbrain. The spatial normalization results were visually checked to assess the quality of the normalization procedure (i.e., anatomical consistency among the registered images).

The intensity normalization procedure consisted of rescaling the voxel values to zero mean and unit standard deviation, making it easier to learn the weights to the optimization algorithm (Raschka and Mirjalili, 2017).

We obtained two preprocessed datasets, detailed as follows, by applying different combinations of spatial and intensity normalization procedures:

- T1-w images processed for intensity normalization (denoted as raw T1-w images) and
- T1-w image obtained by applying spatial and intensity normalization routines (denoted as spatially normalized T1-w images).

Subsequently, we randomly separated the dataset into three subsets: a training set (108 subjects, 60% of the dataset), a validation set (36 subjects, 20% of the dataset), and a test set (36 subjects, 20% of the dataset). We paid particular attention to obtaining the same number of samples per class (i.e., NC and AD) in each subset to get a balanced training dataset.

t-Masking CNN Model

Let X be the training dataset of spatially normalized brain images with $2N$ examples; for a given binary classification task, we define X_0 as the subset of 0-labeled images and X_1 as the subset of 1-labeled images. X was built as a balanced dataset; i.e., X_0 and X_1 have the same size N. Let $X_{0j}(\upsilon)$ be the intensity value of the jth image in X_0 at the voxel υ. The voxel intensity sets $\{X_{0j}(\upsilon) \mid \forall\, j \in [1, N]\}$ and $\{X_{1j}(\upsilon) \mid \forall\, j \in [1, N]\}$ are assumed to refer to the same brain location for all elements since the images are spatially normalized. Let $\mu_0(\upsilon)$ and $\mu_1(\upsilon)$ be the means for the two precedent sets, respectively; hence, we recall the definition of the Student t-statistic (Snedecor and Cochran, 1989), to define the t-map in terms of the t-score:

$$t(v) = \frac{|\mu_0(v) - \mu_1(v)|}{\sqrt{0.5\,(\sigma_0^2(v) + \sigma_1^2(v))}} \tag{1}$$

where σ_0 and σ_1 are the standard deviations for $X_{0j}(\upsilon)$ and $X_{1k}(\upsilon)$ for all j and k in $[1, \ldots N]$, respectively. Considering the null hypothesis H_0: $\mu_0 = \mu_1$, we reject H_0 if $t(v) >$ threshold. Repeating the computation for all v, we obtain a multiple-hypothesis test-based map, namely, the t-mask, corresponding to voxels belonging to brain areas with statistically different values between μ_0 and μ_1, at a given threshold. As a result, only the most salient voxels for the given binary classification task will be considered as relevant features with the t-masking application. Features with 0-value in the t-mask were considered as redundant features, according to the definition of Vergara and Estevez (2014).

The *t*-masking approach was included in a CNN model. Below, we briefly describe the CNN architecture.

Convolutional layers in a 3D CNN are composed of K different filters. Each filter works by convolving an input tensor x with a 3D kernel of weights $W \in R^{m \times m \times m}$ with size $M = m^3$ and adding a bias term b. The result is passed to a non-linear activation function $f(\cdot)$. Therefore, each filter k returns a feature map h_k, extracting a derivative of x, as follows:

$$h_k = f(W_k * x + b_k) \tag{2}$$

The set of K feature maps, extracted from the input x, defines a single layer ℓ in a CNN architecture composed of L layers. The kth feature map at layer ℓ, denoted as $h^l{}_k$, is constructed using the outputs of layer ℓ - 1 as inputs to layer ℓ:

$$h^l_k = f(W^l_k * h^{l-1} + b^l_k) \tag{3}$$

For simplicity, we neglect the description of the pooling layers, which, although fundamental for the CNN model, are not relevant for our purpose. Therefore, the last layer's output θ is inputted in a cross-entropy cost function $J(\theta)$, which depends on a chain of the precedent layers' output (Eq. 3). After a random initialization of the weights, the network is trained in order to update the vector of all weights W and b with gradient-based optimization, using the well-known backpropagation algorithm.

t-Masking induces a simplification of the complex structure of $J(\theta)$. To explain this, we consider the affine transformations of the first layer as

$$W^0_k * x + b^0_k \tag{4}$$

Eq. (4) could be dominated by redundant voxel intensity values. Indeed, Eq. (4) could be reformulated as

$$W^0_k * (x^{rel} + x^{red}) + b^0_k \tag{5}$$

where x^{rel} and x^{red} represent the matrix of relevant and redundant voxel values, respectively, associable to w_i with an appropriate factorization. When $||x^{rel}||_1 \gg ||x^{red}||_1$, the affine transformations are dominated by the redundant terms, principally characterized by low-ranking *t*-scores. Therefore, it should produce a noise-like effect in deeper layers due to relevant patterns hidden by indistinguishable redundant terms. t-Masking imposes $x^{red} = 0$ for all i, improving filters' response when complex interactions in deep layers between x^{rel} and x^{red} can be neglected. Furthermore, it is well known that CNNs have sparse interactions between input features in deep layers (Goodfellow et al., 2017), as evident from Eq. (3). Therefore, considering that *t*-masking preserves the spatial relationship of data, its application to a CNN allows sparse interactions among all relevant voxels ($x_{ij}{}^{rel}$, thus serving the requirements of a CNN. Consequently, the model can extract meaningful information just from the relevant voxels' spatial distribution. This could be an advantage over applying *t*-masking to a classical ML model, which generally requires the vectorization of the input features.

In conclusion, such formal description highlights how *t*-masking can minimize the redundancy, reducing the features' dimensionality, thus removing possible bias sources in the classification task.

Feature Selection

Figure 1 shows a synthetic scheme of the learning procedure. We performed a voxel-wise two-sample *t*-test between spatially normalized T1-w images from NC and very mild AD, with Gaussian smoothing with a full width at half maximum of 6 mm. We applied the *t*-masking to all data (training, validation, and test set) as a binary mask (with a fixed threshold).

We computed the *t*-map, $t(v)$ only on the training set to avoid leakage, i.e., the creation and usage of variables (like labels in the case of classification tasks) that carry information about the outcome of the classification task (Mateos-Pérez et al., 2018). The choice of a suitable threshold for the previous step was investigated, evaluating the classification performance at variable *t*-masking threshold through validation curves. We used the accuracy and the area under the curve (AUC) of receiver operating characteristic (ROC) as metrics for the plots. Train, validation, and test accuracy, along with validation and test AUC curves, are plotted (**Figure 4**) to give a qualitative assessment of the bias, variance, and the model's stability as the threshold changes.

CNN Architecture Overview

We utilized a 3D CNN inspired by the study of Spasov et al. (2019) that already proposed an efficient CNN model (about 5,105 parameters) with high accuracy for AD/MCI conversion prediction. They implemented a dual-learning approach combining several input streams such as structural MRI measures, Jacobian determinant images, and clinical data. We modified such architecture by simplifying and tailoring it for our purposes. In particular, with a Siamese approach, input data were kept only from MRI, eliminating clinical data input, their associated subnetwork, and Jacobian determinant images. Moreover, we removed the dual-learning method, preserving a logistic regression as a binary classifier. The network was designed to receive input masked MR images on two parallel layers. After two layers, the outputs were concatenated, merging the activation maps along the channel axis (the concatenate layer). The add-block performed element-wise addition between two sets of activation maps of the same size along all dimensions. A residual connection in the add-block facilitated training in analogy to ResNet behavior. The network decreased the image inputs' dimensionality using standard, separable, and grouped convolutional blocks before two fully connected layers. Its output was a four-dimensional feature vector, ready for the logistic regression classifier. The following network settings remained unchanged in all experiments: (i) the dropout rate, set at 0.1 for all layers and blocks; (ii) the L2 regularization penalty coefficient, set at 5×10^{-5} for all parameters in the convolutional and fully connected layers. The convolutional kernel weight initialization followed the procedure described by He et al. (2015). The

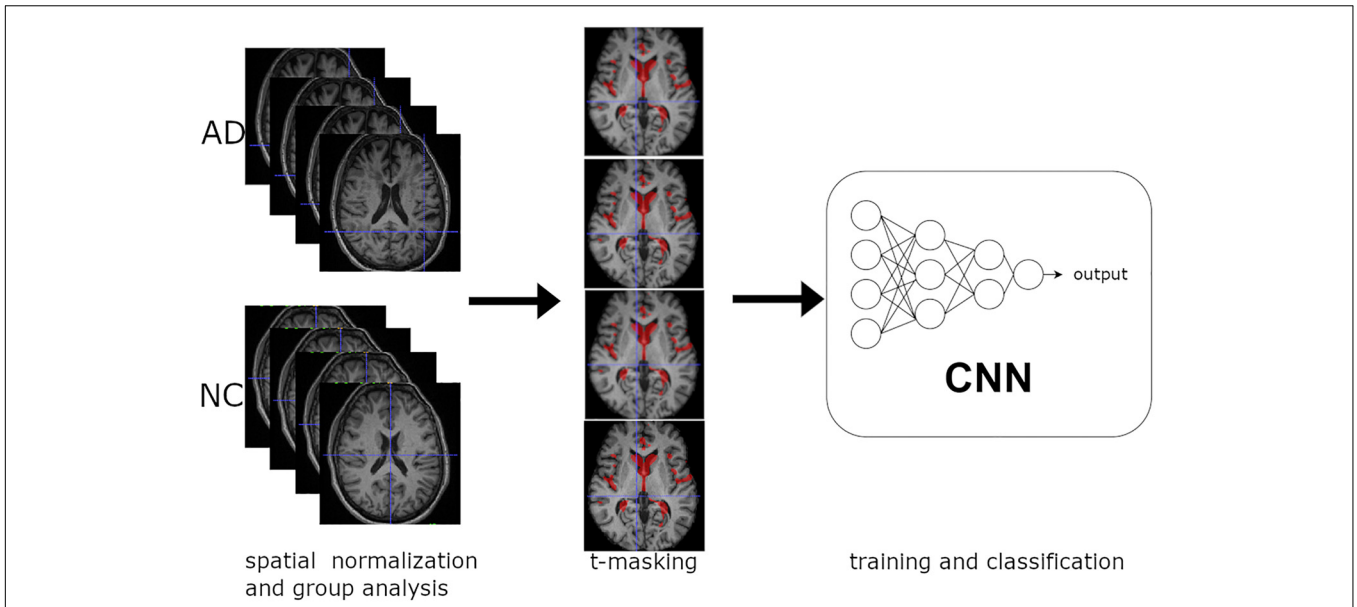


FIGURE 1 | Overview of the learning procedure: (1) Spatial normalization of MR structural images and group analysis on training data. (2) *t*-Masking of input images. (3) Training and classification of 3D CNN.

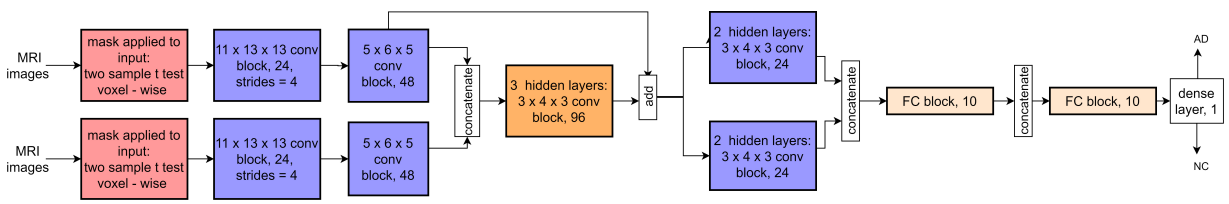


FIGURE 2 | The architecture of our 3D CNN with a Siamese approach. The notation of blocks is as follows: kernel size, (separable) convolutional block, and output channels. If the strides are different from the default value of 1, the new stride value is shown in addition at the end. Each (separable) convolutional block sequentially contains a 3D (separable) convolutional layer, batch normalization, ELU activation, 3D max-pooling (only the convolutional layer), and dropout. FC blocks contain a fully connected layer, batch normalization, ELU activation, and dropout. For details, see Spasov et al. (2019).

objective function loss was minimized using the Adam optimizer (Kingma and Ba, 2017), with an exponentially decaying learning rate of 0.001 × 0.3 epoch/10. A training batch size of six samples was randomly sampled from the dataset when training the network until the dataset was exhausted.

A synthetic scheme showing the adopted CNN architecture is presented in **Figure 2**.

Regularization and Hyperparameter Choice

We adopted several strategies to manage overfitting, some of which were in common with the Spasov model (batch normalization, dropout, and L2 regularization). Since the previous work of Spasov already optimized the hyperparameters like batch size, learning rate, dropout rate, and the L2 regularization parameter, we used the proposed configuration for those parameters. To control the overfitting and reduce computational time, a further implicit regularization, the early stopping (Yao et al., 2007), was included. It consists of stopping the training of the network before it ceases to improve generalization performance. In particular, we imposed the patience parameter at 10 epochs; i.e., we retrieved the model with the minimum validation loss if the last 10 epochs did not obtain a lower validation loss.

Experimental Models

To assess the classification performance of the proposed CNN model, we implemented seven different experimental configurations, as detailed in the following points:

1. 3D CNN model, without FS, on raw T1-w images (raw_MRI).
2. The same model as point 1 to spatially normalized T1-w images (norm_MRI).
3. 3D CNN model with *t*-masking FS (fs_CNN_MRI).
4. A linear classifier with an FS based on *t*-masking (fs_linear_MRI).
5. 3D CNN model with *t*-masking FS corrupted by adding Gaussian noise with 0 mean and 0.2 standard deviation (fs_noise_CNN_MRI).

6. 3D CNN model with *t*-masking FS biased by randomly turning off voxels in the *t*-mask, following a Bernoulli probability distribution with $p = 0.05$ (fs_bernoulli_CNN_MRI).
7. 3D CNN model with a random FS based on a "voxel-wise" dropout that turns off the voxels with a constant Bernoulli distribution probability p. A validation curve was performed as a function of p ranging from 0.01 to 0.99 (fs_random_CNN_MRI).

We implemented fs_linear_MRI to verify the classification performance of the voxel-wise two-sample *t*-test without the CNN. In that experimental setup, we replaced the CNN with a linear classifier that receives in input the mean of the masked image voxel values. The application of *t*-masking to the linear decision model (classification of average intensity voxels by a perceptron) provided an estimate of the statistical mapping contribution alone.

The fs_noise_CNN_MRI and fs_bernoulli_CNN_MRI processes allowed us to assess the robustness of fs_CNN_MRI to small variations of selected voxels by perturbing the *t*-mask. fs_noise_CNN_MRI was thought to assess the *t*-masking's robustness to structural MRI misalignment and subjects' misclassifications. fs_bernoulli_CNN_MRI was planned to evaluate the robustness to type I and II error variations. We implemented fs_random_CNN_MRI to compare *t*-masking applied to the same CNN using another FS technique.

Each experimental point was obtained by averaging metrics acquired from five random (i.e., with a random seed) sample permutations. Raw_MRI and norm_MRI correspond to a single experimental point. In all other cases, we evaluated a validation curve from the applied experimental points to study the different behaviors at different threshold values of the *t*-map. In detail, for models from 3 to 6, we evaluated 21 experimental points, corresponding to several threshold points related to the binary mask obtained from the *t*-map. The threshold values were selected in the equally spaced range [0, 8] with a zero-threshold corresponding to the entire brain mask and an eight-threshold representing no surviving voxels in the binary mask.

Performance Evaluation

We compared the computed models' performance metrics (from model 3. to 6. listed above) through validation curves as a function of the *t*-mask threshold. Test accuracy and ROC analysis were executed by computing the AUC. We adopted a dedicated cross-validation method to determine the quantitative effect of random sampling for splitting the dataset on the training process. In particular, we tested whether the FS method outperformed the baseline CNN by considering the average between each permutation of test accuracy and test AUC, named the average cross-validation. Instead, in the classical cross-validation, the performance was evaluated by observing test accuracy/AUC corresponding to the model with the highest value of the metrics on the validation set. This method allowed us to identify what thresholds outperform baseline models, and it enabled us to compare the validation curves of all our models in a statistically meaningful way. Moreover, to allow a further comparison with other literature methods, we added the test accuracy and the test AUC obtained by the classically fivefold cross-validation; i.e., we selected the model which achieves the best performance on the validation set.

We evaluated the performance enhancement (PE) of *t*-masking by comparing the peak accuracy of fs_CNN_MRI with the accuracy of the raw_MRI model. In particular, PE is computed as the difference of the test accuracy of two models, estimated with average cross-validation; negative PE indicates that the considered model underperforms the reference model. We also compared the PE result with the PE of typical FS approaches compared on a previous work (Tohka et al., 2016). These models are trained on 200 subjects from ADNI MRI T1-w to resolve an AD-vs.-NC classification task (with no CDR subjects' selection).

Furthermore, we included a visual explanation method as a qualitative performance evaluation, based on the computation of a saliency map (heat-map) that localizes relevant image regions for the CNN model, resulting in a saliency map of features. In particular, we implemented a class activation mapping (CAM) method for 3D CNNs, by using the grad-CAM algorithm developed by Selvaraju et al. (2020). The heat-map was derived by computing the gradient between a convolutional layer output and the loss output. We inspected the output of the first "concatenate" layer (**Figure 2**), as a compromise between the deepest layer possible for maintaining high-level semantics and the highest map resolution (Selvaraju et al., 2020). We applied grad-CAM to the models derived from each considered threshold of fs_CNN_MRI and norm_MRI.

Algorithm Implementation

All experiments were conducted using Python 3.6.9. The neural network was built with the Keras DL library using TensorFlow 2.0 as a backend. 3D convolutions were available as a Keras module. The 3D separable and grouped convolutions modules were available in previous works (Spasov et al., 2019). The code, available to https://github.com/simeon-spasov/MCI, was adapted for the v.2 of TensorFlow[1] and v.3 of Python. To compute the *t*-map, we employed the Python module nistats, included in nilearn ("Nistats: Functional MRI in Python—functional MRI for NeuroImaging," n.d.). For the whole experimental procedure, we utilized a cloud computing strategy: a free Jupiter notebook environment working directly on a browser with a virtual machine that requires no setup to use while providing free access to computing resources, including GPU (up to 25.51-GB RAM, a GPU T4 or P100). The code to replicate the experiments is available at https://gitlab.com/sdndeep/voxel_wise_fs.

Employing the resources available for Colab users (Colab offers only one core CPU), we run *t*-masking with limited computational power. Therefore, currently available workstations (or PCs) on the market could obtain better time performances, considering a faster charge dataset on RAM and the statistical test's parallelizability.

It is important to underline that the addition of *t*-masking provides an additional computational burden which may not

[1] https://www.tensorflow.org/api_docs?hl=en

be negligible when large datasets are used. Indeed, it should be considered that to perform the two-sample *t*-test, the entire training set must be loaded in RAM memory to compute the *t*-mask.

RESULTS

In **Figure 3**, we show an example of *t*-masks, obtained within a fs_CNN_MRI training run, at different thresholds of *t*-parameter, in the range [0.0, 6.0].

Table 1 shows the results of the tested experimental models, in terms of both accuracy and AUC.

Overall, the fs_CNN_MRI model holds the highest performance for both average and classical cross-validation metrics. In particular, considering the average cross-validation metric, the fs_CNN_MRI outperforms the raw_MRI and norm_MRI models with PE values of (6 ± 2)% and (7 ± 2)% (accuracy) and differences in AUC of 0.06 ± 0.02 and 0.05 ± 0.02, respectively. The results were also confirmed for classical cross-validation metrics, where fs_CNN_MRI outperforms both the raw_MRI (PE: 3%; AUC difference: 0.06) and norm_MRI (PE: 3%; AUC difference: 0.05).

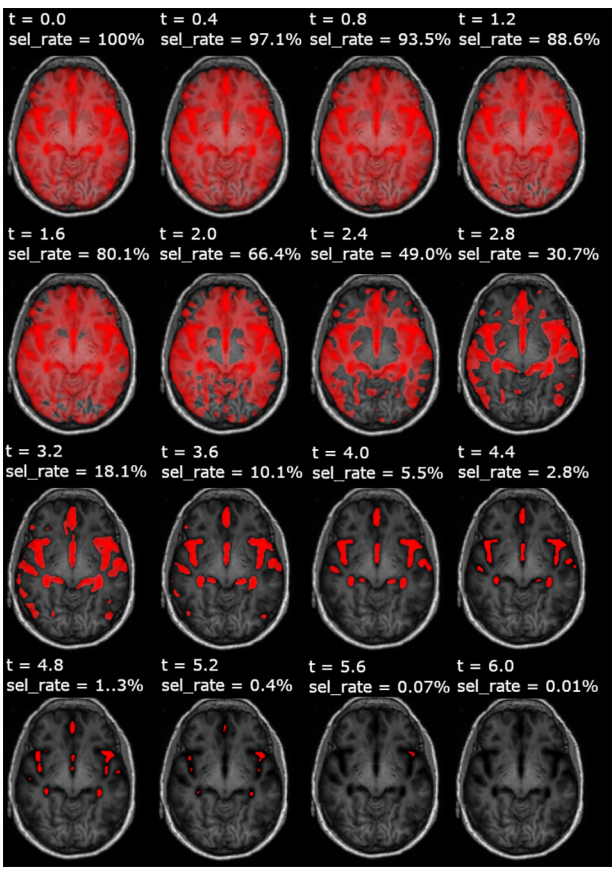


FIGURE 3 | Feature selection maps (*t*-masking) overlaid on a spatially normalized axial T1-w images, for each adopted threshold (*t*). This figure represents an example resulting from a *t*-masking pipeline run applied to very mild AD- vs. -NC classification. For each *t*-masking, the fraction of voxels selected for a given threshold with respect to the whole-brain volume is also reported (sel_rate).

Figure 4 shows the *t*-mask threshold setting's influence on the performance of the CNN models in terms of validation curves for all metrics (train–validation–test accuracy, validation, and test AUC). We observe that the average performance increases until reaching a peak in all models utilizing the FS method. For all *t*-masking models, the curves show how the metrics reach a peak value, improving the performance to the threshold increase until a final breakdown due to the loss of relevant voxels. The peak is in the same threshold range [3, 5] for all models, which means, recalling **Figure 3**, the percentage of selected voxels is between approximately 18% and 0.4%.

We evaluated the PE of *t*-masking using average cross-validation results for test accuracy between fs_CNN_MRI and raw_MRI. We obtained a PE of (6 ± 2)%, as shown in **Table 1**. **Table 2** shows a comparison of our experimental PE results with a precedent work in the literature (Tohka et al., 2016). The original article's notation is maintained. In this paper, the authors studied an AD-vs.-NC classification task, applying several FS techniques to classical ML classifiers. They compared SVM, with or without filter-based FS; several embedded FS methods; and stability selection (with logistic regression, lasso, elastic net, and graph net). Filter-based methods consisted of a *t*-test-based filter, with or without a false discovery rate (FDR)-corrected threshold. Their classification task is not directly comparable since they used the ADNI dataset without a CDR selection.

Finally, to provide the reader with an idea of the operational feasibility of the method, we evaluated the additional computational time required by *t*-masking; it took about 5 min for the training dataset (108 subjects), considering the experimental setup described above.

The heat-maps derived from the grad-CAM algorithm, at varying *t*-thresholds for the fs_CNN_MRI experiment, are reported in **Supplementary Figure 1**. It is clear, from the heat-maps, that the greater the weight of the *t*-masking, the greater the focus of the grad-CAM model on the salient regions of interest for the classification task.

DISCUSSION

We proposed a *t*-masking FS method for deep neural networks on a CNN architecture by evaluating its performance and the behavior with respect to the adopted salience parameter (*t*-threshold). We tested the proposed CNN model for a very mild AD-vs.-NC classification task with an structural MRI dataset of 180 subjects. It is the first time a feature selection technique is investigated for DL applications to neuroimaging, to the best of our knowledge. By comparing raw_MRI and norm_MRI models with fs_CNN_MRI, we can affirm that the adoption of the *t*-masking resulted in an effective increase in performance. In particular, it should be noted that raw_MRI and norm_MRI reached similar performance. For this reason, we can deduce that

TABLE 1 | Test accuracy and AUC evaluated for each tested method with average cross-validation (the average between permutations) and classical cross-validation (the highest between permutations).

Experimental models	Average cross-validation		Classical cross-validation	
	Test accuracy	Test AUC	Test accuracy	Test AUC
raw_MRI	(69 ± 2)%	0.79 ± 0.01	78%	0.82
norm_MRI	(68 ± 2)%	0.80 ± 0.01	78%	0.83
fs_CNN_MRI (*t* = 3.6)	(75 ± 1)%	0.85 ± 0.01	81%	0.88
fs_linear_MRI (*t* = 4.4)	(69 ± 2)%	0.78 ± 0.02	72%	0.81
fs_noise_CNN_MRI (*t* = 4.4)	(76 ± 3)%	0.84 ± 0.03	72%	0.81
fs_bernoulli_CNN_MRI (*t* = 4.4)	(74 ± 3)%	0.82 ± 0.03	69%	0.83
fs_random_CNN_MRI (*p* = 71%)	(57 ± 8)%	0.62 ± 0.15	64%	0.69

Please refer to the text for the details of the tested methods.

the spatial normalization process did not degrade the classification performance and, notably, that the enhancement of fs_CNN_MRI can be reasonably attributed solely to the *t*-masking.

Moreover, the similar performance between fs_linear_MRI (which applies *t*-masking to a linear model) and both raw_MRI and norm_MRI indicated that the classification based only on statistical mapping achieved similar results compared with using CNN alone, whereas it reached worse results compared with fs_CNN_MRI. We showed that in fs_noise_CNN_MRI and fs_bernoulli_CNN_MRI, the overfitting slightly increases. We should also note that validation curves verified the increasing linear dependence of the performance as a function of the selected threshold (for models from 3. to 6.). It is interesting that the PE peak was reached in the same threshold range [3, 5] for all *t*-masking models. This observation can be interpreted as *t*-masking results being robust to small attempts to mess up the *t*-map, despite a slight increase in overfitting and inferior generalization capability. Finally, fs_random_CNN_MRI model results showed that the performance, in terms of both accuracy and AUC, decreased if a random voxel-wise FS replaces the *t*-mask. This comparison is a further confirmation of how *t*-masking selected relevant features against the random FS approach. As shown in **Figures 3**, **4**, relevant features, i.e., features leading to better classification performance, represented a small 10.1% fraction of the total brain, thus confirming the working hypothesis explained in Section 2.3 ($||x^{rel}||_1 \gg ||x^{red}||_1$). Furthermore, grad-CAM results show that *t*-masking contributes to the focalization of relevant image regions for the model training with respect to norm_MRI ($t = 0.0$), which instead shows a more dispersive feature distribution. As expected, as the *t*-threshold grows, the heat-maps highlight more focused regions for the classification task.

We analyzed early AD vs. NC, as a test bench, for a demonstrated difference between AD brains and NC brains, due to the presence of subtle region-specific anatomical alterations in AD brains, viewable via structural MRI T1-w and detectable by a voxel-wise *t*-test. The obtained results justify this choice and allow us to deduce that the same method could be applied to disorders with the same characteristics.

In the method here presented, we have chosen a voxel-wise two-sample *t*-test, considering that it is a standard method for mass univariate analysis and represents a well-established statistical tool for group analysis in the neuroimaging field. In principle, other statistical tests, such as mass multivariate methods, could serve as suitable FS methods and deserve further investigation. Indeed, considering **Figure 3**, other statistical maps could obtain a similar distribution of relevant voxels.

In this work, the CNN architecture proposed by Spasov et al. (2019) was adopted due to both its high performance and efficiency in the AD-vs.-NC classification task. However, different CNN architectures could be used, and further investigations are required to assess the *t*-masking performance when different CNN architectures are adopted.

We obtained the best PE with a reduction of 90% brain voxels, protecting from the curse of dimensionality. However, in the DL literature, FS techniques are often under-investigated since deep neural networks aim to implicitly extract the relevant features. Based on our findings, we argued that FS techniques, like *t*-masking, should deserve more attention in DL models' design. The *t*-masking approach is integrated into the learning process and is based only on simple statistical analysis on training data, with limited *a priori* hypotheses. For this reason, it is in line with the fundamental ideas of DL.

It is possible to adopt data augmentation, pre-trained layers, or transfer learning strategies to obtain better performance. Transfer learning reaches state-of-the-art performance in pathological brain detection with AlexNet-based neural network architectures, as demonstrated by Lu et al. (2020) and Lu et al. (2019). We excluded the application of these techniques since this work was focused just on the impact of FS on the performance. However, the study of the impact of the concurrent combination of multiple strategies to the global performance of a DL model deserves attention and further investigations.

In the context of the CDR-based AD-vs.-NC classification task with transfer learning, Nanni et al. (2020) reached 0.83 AUC when only very mild AD subjects (CDR = 0.5) were selected. With a 0.88 test AUC, the proposed *t*-masking model fs_CNN_MRI outperformed the ensemble transfer learning model for the same classification task and reached comparable performance when a conventional ML was applied to the same classification task (0.89 AUC).

Our results show that *t*-masking could not improve the performance under some specific conditions. Indeed, we have

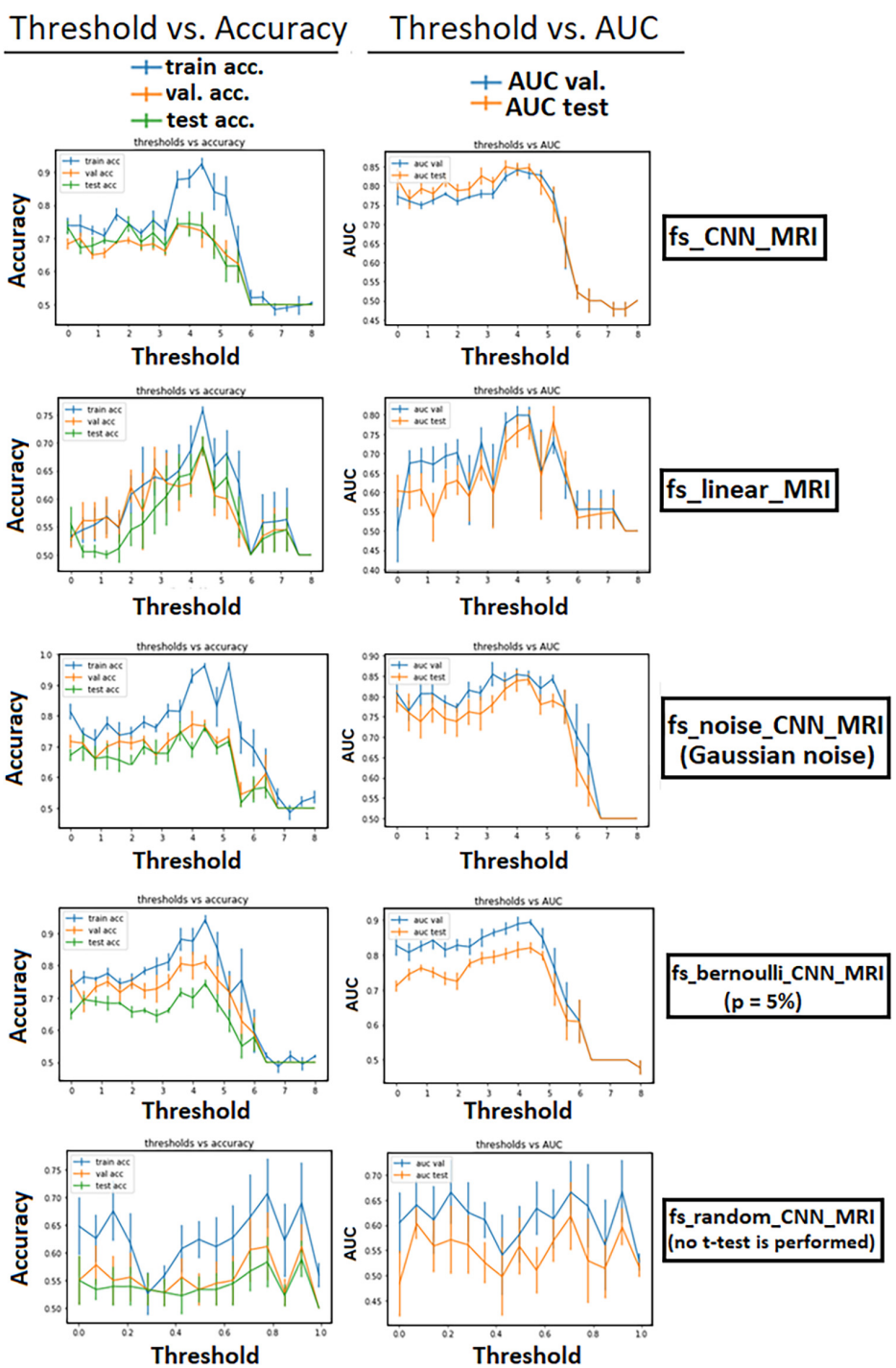


FIGURE 4 | Performance of the models as a function of the *t*-mask threshold. Error bars represent the standard deviation of the mean. For all *t*-masking models, the curves show how the metrics reach a peak value, improving the performance to the threshold increase until a final breakdown due to the loss of relevant voxels. The zero-point thresholds are equivalent measures of the norm_MRI model for fs_CNN_MRI and fs_noise_CNN_MRI. The bottom row shows the results in the case of "voxel-wise" dropout (with increasing *p*-value on the abscissa). There is evidence that the performance of fs_random_CNN_MRI rapidly decreases at increasing *p*-value.

found that when the number of selected voxels was limited, the performance deteriorated with increasing overfitting. This result could be ascribed to the loss of relevant information when decreasing the number of voxels. If *t*-masking is considered for different classification tasks, the active regions (i.e., regions in which voxels are selected by masking) might be too few. A high

TABLE 2 | Comparison of PE for models using FS techniques, analyzed in Tohka et al. (2016).

	Performance enhancement
EN-VACV	+1.8%
EN-VABEE	+1.4%
EN-05CV	+1.8%
EN-05BEE	+1.0%
LASSOCV	+1.4%
LASSOBEE	0%
LASSOSTAB	−0.2%
EN-05STAB	−0.1%
GNCV	+1.5%
GNBEE	+0.7%
SVMF-FDR	+1.3%
SVMF-1000	+1.6%
SVMF-125	+2.1%
SVM-ALL	0%
***t*-Masking CNN**	**+6 ± 2%**

The original article's notation is maintained. The results are referred to an average of the experimental conditions with resampled images to 4 and 8 mm isotropic and spatial resolutions. In the case of results from Tohka et al. (2016), the PE is computed by subtracting the test accuracy of the model under consideration with SVM-ALL test accuracy; in the case of fs_CNN_MRI, the PE is computed by a subtraction with raw_MRI. Negative values refer to worse performance.

variability as a function of the training set could be found, depending on the brain disorder's intrinsic physiology. A possible solution could be to use different imaging modalities or to exploit multimodal imaging. A further limitation of our study is represented by the dataset size, and further investigations are required to test how our FS technique improves the classification performance capabilities on a larger dataset.

The FS *t*-masking approach deserves attention also for the possibility of being easily generalizable for the classification of other pathologies, as well as for other neuroimaging modalities (if it is possible to make a voxel-wise comparison through a spatial normalization). Furthermore, the *t*-masking, although demonstrated on a CNN in this work, can, in principle, be used to perform FS in other DL architectures.

Considering the comparison with results obtained by Tohka et al. (2016), as reported in **Table 2**, it should be noted that their experimental procedure differed from ours due to the use of a split-half resampling-type analysis. However, they reported an average test accuracy. For this reason, average cross-validation results are more suitable for comparison than classical cross-validation, despite the latter method being more used in the literature. Despite the less challenging classification task (AD vs. NC with respect to very mild AD vs. NC, in our work) and the differences in the analyzed dataset (for example, MRI T1-w images acquired with a 1.5-T scanner), it is interesting to compare their results with ours by using the PE as a more objective comparison criterion. In particular, we have achieved a threefold improvement in performance compared to their algorithms. It is useful to underline that the model SVM-ALL refers to an SVM classifier without any FS application and, consequently, that it is possible to compare the PE obtained by Tohka et al. (2016) by using FS methods. Accuracy comparison cannot be applied due to the different data sources. Moreover, the presented comparison should be analyzed considering that they did not select AD subjects using a CDR-based analysis. We did not find other studies analyzing a very mild AD-vs.-NC binary classification task on the OASIS-3 dataset. However, Saraswathi et al. (2013) used a GA-ELM-PSO (refer to the article for the notation) classifier for AD multiclass classification by CDR-level voxel-based morphometry for feature extraction and a genetic algorithm for FS. They used an OASIS dataset (not OASIS-3), obtaining an 81% test accuracy for the very mild classification. In real-world applications, *t*-masking could be suitable in contexts with hundreds (or few thousands) of subjects available for training. The use of *t*-masking indeed requires increased use of resources during training. In particular, the computation of the *t*-mask requires an additional computational step in the training phase, thus requiring proper RAM resources and extra computational time. Therefore, the *t*-masking CNN model pays the performance improvement with a higher training time than a CNN alone because *t*-masking takes about 5 min to compute the *t*-mask of 108 subjects. Besides, if properly managed, the voxel selection operated by *t*-masking can, in principle, shorten the training time due to sparse active voxels and improve the DL model interpretability.

CONCLUSION

In this work, we analyzed the application of a group-analysis-based FS, the *t*-masking, to deep neural network architecture, in the case of a very mild AD-vs.-NC classification task on a 180-example dataset of structural MR. We showed that the *t*-masking application could enhance the classification performance even better than could conventional FS techniques. Moreover, *t*-masking is generalizable to other binary classification tasks, different neuroimaging modalities, and other DL architectures.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.oasis-brains.org (OASIS-3).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comitato Etico IRCCS Pascale and OASIS-3 data sharing was granted by participants through informed consent and local IRB approval. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DM: conceptualization, methodology, investigation, and writing–original draft. PB: methodology, writing, reviewing, and editing. PR and MS: writing, reviewing, editing, and supervision. MA: conceptualization, investigation, writing, reviewing, editing,

and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

Data were provided by OASIS-3 (principal investigators: T. Benzinger, D. Marcus, and J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, and R01EB009352). This work was partially funded by the Italian Ministry of Health ("ricerca corrente" and "GR-2018-12366779" projects) and by the Italian Ministry for Education, University and Research (MIUR) for the project "Metodi innovativi di imaging molecolare per lo studio di malattie oncologiche e neurodegenerative–MOLIM ONCOBRAIN LAB"(ARS01 00144 Prot. U.0013785 31-08-2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.630747/full#supplementary-material

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* 8:14. doi: 10.3389/fninf.2014.00014

Afzal, S., Maqsood, M., Nazir, F., Khan, U., Aadil, F., Awan, K. M., et al. (2019). a data augmentation-based framework to handle class imbalance problem for Alzheimer's stage detection. *IEEE Access* 7, 115528–115539. doi: 10.1109/ACCESS.2019.2932786

Aiello, M., Cavaliere, C., D'Albore, A., and Salvatore, M. (2019). The challenges of diagnostic imaging in the era of big data. *J. Clin. Med.* 8:316. doi: 10.3390/jcm8030316

Asami, T., Hyuk Lee, S., Bouix, S., Rathi, Y., Whitford, T. J., Niznikiewicz, M., et al. (2014). Cerebral white matter abnormalities and their associations with negative but not positive symptoms of schizophrenia. *Psychiatry Res.* 222, 52–59. doi: 10.1016/j.pscychresns.2014.02.007

Asim, Y., Raza, B., Malik, A. K., Rathore, S., Hussain, L., and Iftikhar, M. A. (2018). A multi-modal, multi-atlas-based approach for Alzheimer detection via machine learning. *Int. J. Imaging Syst. Technol.* 28, 113–123. doi: 10.1002/ima.22263

Audoin, B., Zaaraoui, W., Reuter, F., Rico, A., Malikova, I., Confort-Gouny, S., et al. (2010). Atrophy mainly affects the limbic system and the deep grey matter at the first stage of multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 81, 690–695. doi: 10.1136/jnnp.2009.188748

Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., et al. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur. J. Cancer* 113, 47–54. doi: 10.1016/j.ejca.2019.04.001

Chatterjee, A., Vallières, M., and Seuntjens, J. (2020). Overlooked pitfalls in multi-class machine learning classification in radiation oncology and how to avoid them. *Phys. Med.* 70, 96–100. doi: 10.1016/j.ejmp.2020.01.009

Chaves, R., Ramírez, J., Górriz, J. M., López, M., Salas-Gonzalez, D., Álvarez, I., et al. (2009). SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting. *Neurosci. Lett.* 461, 293–297. doi: 10.1016/j.neulet.2009.06.052

Chen, Z., Pang, M., Zhao, Z., Li, S., Miao, R., Zhang, Y., et al. (2020). Feature selection may improve deep neural networks for the bioinformatics problems. *Bioinformatics* 36, 1542–1552. doi: 10.1093/bioinformatics/btz763

Choi, H., Jin, K. H., and Alzheimer's Disease Neuroimaging Initiative (2018). Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav. Brain Res.* 344, 103–109. doi: 10.1016/j.bbr.2018.02.017

Davatzikos, C. (2019). Machine learning in neuroimaging: progress and challenges. *Neuroimage* 197, 652–656. doi: 10.1016/j.neuroimage.2018.10.003

Ding, Y., Sohn, J. H., Kawczynski, M. G., Trivedi, H., Harnish, R., Jenkins, N. W., et al. (2019). A deep learning model to predict a diagnosis of Alzheimer Disease by using 18F-FDG PET of the brain. *Radiology* 290, 456–464. doi: 10.1148/radiol.2018180958

Dora, L., Agrawal, S., Panda, R., and Abraham, A. (2017). State-of-the-art methods for brain tissue segmentation: a review. *IEEE Rev. Biomed. Eng.* 10, 235–249. doi: 10.1109/RBME.2017.2715350

Fan, J., and Fan, Y. (2008). High-dimensional classification using features annealed independence rules. *Ann. Statist.* 36, 2605–2637. doi: 10.1214/07-AOS504

Fan, Y., Shen, D., Gur, R. C., Gur, R. E., and Davatzikos, C. (2007). COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Transac. Med. Imaging* 26, 93–105. doi: 10.1109/TMI.2006.886812

Fox, N. C., and Freeborough, P. A. (1997). Brain atrophy progression measured from registered serial MRI: validation and application to alzheimer\textquotesingles disease. *J. Magn. Reson. Imaging* 7, 1069–1075. doi: 10.1002/jmri.1880070620

Goodfellow, I., Bengio, Y., and Courville, A. (2017). *Deep Learning*. Cambridge, MA: Mit Pr.

Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., and Rueckert, D. (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage* 65, 167–175. doi: 10.1016/j.neuroimage.2012.09.065

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 1026–1034. doi: 10.1109/ICCV.2015.123

Ijaz, M. F., Alfian, G., Syafrudin, M., and Rhee, J. (2018). Hybrid prediction model for type 2 diabetes and hypertension using DBSCAN-Based outlier detection, synthetic minority over sampling technique (SMOTE), and random forest. *Appl. Sci.* 8:1325. doi: 10.3390/app8081325

Ijaz, M. F., Attique, M., and Son, Y. (2020). Data-driven cervical cancer prediction model with outlier detection and over-sampling methods. *Sensors* 20:2809. doi: 10.3390/s20102809

Jo, T., Nho, K., and Saykin, A. J. (2019). Deep learning in Alzheimer's Disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front. Aging Neurosci.* 11:220. doi: 10.3389/fnagi.2019.00220

Khagi, B., Lee, B., Pyun, J.-Y., and Kwon, G.-R. (2019). "CNN models performance analysis on MRI images of OASIS dataset for distinction between healthy and Alzheimer's patient," in *in 2019 International Conference on Electronics, Information, and Communication (ICEIC)*, Auckland, 1–4. doi: 10.23919/ELINFOCOM.2019.8706339

Kingma, D. P., and Ba, J. (2017). Adam: a method for stochastic optimization. *arXiv*. [Preprint] Available online at: http://arxiv.org/abs/1412.6980 (accessed September 28, 2020).

LaMontagne, P. J., Benzinger, T. L. S., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., et al. (2019). OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer Disease. *medRxiv* [Preprint]. doi: 10.1101/2019.12.13.19014902

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

Ling, C. X., Huang, J., and Zhang, H. (2003). "AUC: a better measure than accuracy in comparing learning algorithms," in *Advances in Artificial Intelligence Lecture Notes in Computer Science*, eds Y. Xiang and B. Chaib-draa (Berlin: Springer), 329–341. doi: 10.1007/3-540-44886-1_25

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88. doi: 10.1016/j.media.2017.07.005

Lu, S., Lu, Z., and Zhang, Y. (2019). Pathological brain detection based on AlexNet and transfer learning. *J. Comput. Sci.* 30, 41–47.

Lu, S., Wang, S.-H., and Zhang, Y.-D. (2020). Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm. *Neural Comput. Appl.* doi: 10.1007/s00521-020-05082-4

Lundervold, A. S., and Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* 29, 102–127. doi: 10.1016/j.zemedi.2018.11.002

Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., and Evans, A. C. (2018). Structural neuroimaging as clinical predictor: a review of machine learning applications. *NeuroImage* 20, 506–522. doi: 10.1016/j.nicl.2018.08.019

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., and Tohka, J. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104, 398–412. doi: 10.1016/j.neuroimage.2014.10.002

Mwangi, B., Tian, T. S., and Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinform* 12, 229–244. doi: 10.1007/s12021-013-9204-3

Nanni, L., Interlenghi, M., Brahnam, S., Salvatore, C., Papa, S., Nemni, R., et al. (2020). Comparison of transfer learning and conventional machine learning applied to structural brain mri for the early diagnosis and prognosis of Alzheimer's Disease. *Front. Neurol.* 11:576194. doi: 10.3389/fneur.2020.576194

Nistats (n.d.). *Functional MRI in Python — Functional MRI for NeuroImaging*. Available online at: https://nistats.github.io/index.html (accessed October 2, 2020).

Pardoe, H., Pell, G. S., Abbott, D. F., Berg, A. T., and Jackson, G. D. (2008). Multi-site voxel-based morphometry: methods and a feasibility demonstration with childhood absence epilepsy. *NeuroImage* 42, 611–616. doi: 10.1016/j.neuroimage.2008.05.007

Raj, R. J. S., Shobana, S. J., Pustokhina, I. V., Pustokhin, D. A., Gupta, D., and Shankar, K. (2020). Optimal feature selection-based medical image classification using deep learning model in internet of medical things. *IEEE Access* 8, 58006–58017. doi: 10.1109/ACCESS.2020.2981337

Raschka, S., and Mirjalili, V. (2017). *Python Machine Learning*. Birmingham: Packt Publishing Ltd.

Sage, C. A., Peeters, R. R., Görner, A., Robberecht, W., and Sunaert, S. (2007). Quantitative diffusion tensor imaging in amyotrophic lateral sclerosis. *NeuroImage* 34, 486–499. doi: 10.1016/j.neuroimage.2006.09.025

Salmanpour, M. R. (2020). Machine learning methods for optimal prediction of motor outcome in Parkinson's disease. *Phys. Med.* 8, 233–240.

Saraswathi, S., Mahanand, B. S., Kloczkowski, A., Suresh, S., and Sundararajan, N. (2013). "Detection of onset of Alzheimer's disease from MRI images using a GA-ELM-PSO classifier," in *2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI)*, Singapore, 42–48. doi: 10.1109/CIMI.2013.6583856

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* 128, 336–359. doi: 10.1007/s11263-019-01228-7

Senjem, M. L., Gunter, J. L., Shiung, M. M., Petersen, R. C., and Jack, C. R. (2005). Comparison of different methodological implementations of voxel-based morphometry in neurodegenerative disease. *Neuroimage* 26, 600–608. doi: 10.1016/j.neuroimage.2005.02.005

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

Singh, D. A. A. G., Balamurugan, S. A. A., and Leavline, E. J. (2016). Literature review on feature selection methods for high-dimensional data. *Int. J. Comput. Appl.* 136, 9–17.

Singh, S. P., Wang, L., Gupta, S., Goli, H., Padmanabhan, P., and Gulyás, B. (2020). 3D deep learning on medical images: a review. *Sensors* 20:5097. doi: 10.3390/s20185097

Snedecor, G., and Cochran, W. (1989). *Statistical Methods*, 8th Edn. Hoboken, NJ: Wiley.

Spasov, S., Passamonti, L., Duggento, A., Liò, P., and Toschi, N. (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer\textquotesingles disease. *NeuroImage* 189, 276–287. doi: 10.1016/j.neuroimage.2019.01.031

Suk, H.-I., Lee, S.-W., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2016). Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. *Brain Struct. Funct.* 221, 2569–2587. doi: 10.1007/s00429-015-1059-y

Tohka, J., Moradi, E., Huttunen, H., and Alzheimer's Disease Neuroimaging Initiative (2016). Comparison of feature selection techniques in machine learning for anatomical brain MRI in dementia. *Neuroinformatics* 14, 279–296. doi: 10.1007/s12021-015-9292-3

Traverso, A., Kazmierski, M., Zhovannik, I., Welch, M., Wee, L., Jaffray, D., et al. (2020). Machine learning helps identifying volume-confounding effects in radiomics. *Phys. Med.* 71, 24–30. doi: 10.1016/j.ejmp.2020.02.010

Venkatesh, B., and Anuradha, J. (2019). A review of feature selection and its methods. *Cybern. Inf. Technol.* 19, 3–26. doi: 10.2478/cait-2019-0001

Vergara, J., and Estevez, P. (2014). A review of feature selection methods based on mutual information. *Neural Comput. Appl.* 24, 175–186. doi: 10.1007/s00521-013-1368-0

Vieira, S., Pinaya, W. H. L., and Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. *Neurosci. Biobehav. Rev.* 74, 58–75. doi: 10.1016/j.neubiorev.2017.01.002

Yao, Y., Rosasco, L., and Caponnetto, A. (2007). On early stopping in gradient descent learning. *Constr. Approx.* 26, 289–315. doi: 10.1007/s00365-006-0663-2

Yassin, N. I. R., Omran, S., El Houby, E. M. F., and Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. *Comput. Methods Prog. Biomed.* 156, 25–45. doi: 10.1016/j.cmpb.2017.12.012

Zhang, Y.-D., Wang, S., and Dong, Z. (2014). Classification of Alzheimer Disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Prog. Electromagnet. Res.* 144, 171–184. doi: 10.2528/PIER13121310

Zhu, G., Jiang, B., Tong, L., Xie, Y., Zaharchuk, G., and Wintermark, M. (2019). Applications of deep learning to neuro-imaging techniques. *Front. Neurol.* 10:869. doi: 10.3389/fneur.2019.00869

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 06 May 2021
doi: 10.3389/fnins.2021.685119

A Novel Smart Motor Imagery Intention Human-Computer Interaction Model Using Extreme Learning Machine and EEG Signals

Yi Gu and Lei Hua*

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China

The brain is the central nervous system that governs human activities. However, in modern society, more and more diseases threaten the health of the brain and nerves and spinal cord, making the human brain unable to conduct normal information interaction with the outside world. The rehabilitation training of the brain-computer interface can promote the nerve repair of the sensorimotor cortex in patients with brain diseases. Therefore, the research of brain-computer interface for motor imaging is of great significance for patients with brain diseases to restore motor function. Due to the characteristics of non-stationary, nonlinear, and individual differences of EEG signals, there are still many difficulties in the analysis and classification of EEG signals at this stage. In this study, the Extreme Learning Machine (ELM) model was used to classify motor-imaging EEG signals, identify the user's intention, and control external devices. Considering that single-modal features cannot represent the core information, this study uses a fusion feature that combines temporal and spatial features as the final feature data. The fusion features are input to the trained ELM classifier, and the final classification result is obtained. Two sets of BCI competition data in the BCI competition public database are used to verify the validity of the model. The experimental results show that the ELM model has achieved a classification accuracy of 0.7832 in the classification task of Data Sets IIb, which is higher than other comparison algorithms, and shows universal applicability among different subjects. In addition, the average recognition rate of this model in the Data Sets IIIa classification task reaches 0.8347, which has obvious advantages compared with the comparative classification algorithm. The classification effect is smaller than the classification effect obtained by the champion algorithm of the same project, which has certain reference value.

Keywords: human-computer interaction model, ELM, motor imagery, EEG signals, BCI data set

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Shan Zhong,
Changshu Institute of Technology, China
Yuexin Li,
Hubei University, China

****Correspondence:***
Yi Gu
8202101437@jiangnan.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *24 March 2021*
Accepted: *08 April 2021*
Published: *06 May 2021*

Citation:
Gu Y and Hua L (2021) A Novel Smart Motor Imagery Intention Human-Computer Interaction Model Using Extreme Learning Machine and EEG Signals. Front. Neurosci. 15:685119. doi: 10.3389/fnins.2021.685119

INTRODUCTION

The brain is the central nervous system that governs human activities. However, in modern society, more and more diseases threaten the health of the brain and nerves and spinal cord, making the human brain unable to conduct normal information interaction with the outside world. In recent years, with the rapid development of computer science and people's continuous in-depth exploration of brain science, brain-computer interface (BCI) technology has been attracting

attention as a new type of human-computer interaction system that directly interacts between brain nerves and peripheral devices. The BCI system collects EEG by placing sensors on the surface of the user's scalp or inside the skin. The EEG signal is decoded to determine the intention, and then complete a series of operations such as controlling external equipment. Using BCI technology can help patients with movement disorders caused by nerve damage regain the ability to move independently and smoothly interact with the outside world. This technology will significantly improve the quality of life of patients and reduce the burden on families and society. The first international BCI conference held in 1999 clearly defined the brain-computer interface, that is, "BCI is a communication control system for direct communication between the brain and external devices, and does not rely on brain nerves and peripheral muscle tissue" (Wolpaw et al., 2000). Nowadays, BCI technology is developing rapidly at a rapid pace. In the fields of military aviation, rehabilitation and medical treatment, cognitive enhancement, games and entertainment, and intelligence are bursting with strong vitality, the research and application of brain-computer interface has become a hot spot in the field of scientific research (Van Dokkum et al., 2015; Qiu et al., 2017; Hammer et al., 2018).

Motor Imagery (MI) EEG has the characteristics of flexibility, non-invasiveness, low environmental requirements, and high resolution. Therefore, MI is one of the widely used forms of BCI. The frequency band power of the EEG signal during the motion imaging process will vary with the content of the MI task, which is called event related synchronization/desychronization (ERS/ERD). The generation of ERS/ERD is related to internal or external events. When one side of the human limb exercises or performs motor imagination, the energy of μ rhythm and β rhythm in the sensory motor area on the opposite side of the brain decreases, and the energy of μ rhythm and β rhythm in the ipsilateral motor sensory area increases. This rule makes it possible for ERS/ERD to control external equipment or perform motor imagination intention recognition (Lisi et al., 2014). Motor Imagery Brain-Computer Interface (MI-BCI) is used as a branch of the brain-computer interface. The patient's brain imagines and simulates actions, but there is no actual action output. MI-BCI is mainly based on the analysis and recognition of sensorimotor rhythms, which decodes the signals and converts them into machine instructions. It can establish information channels between humans and machines, and realize the control of devices such as wheelchairs, exoskeletons, and prostheses. According to the theory of neuroplasticity, it also helps to activate nerve cells in a specific area of activity, helping to repair and regenerate damaged nerves.

In 1973, the American Jacques Vidal team put forward the concept of brain-computer interface technology for the first time in a paper, and pointed out that it is an interdisciplinary technology covering various disciplines, including biomedicine, neuroscience, cognitive science, computer science and other fields (Vidal, 1973). However, due to the immaturity of various corresponding technologies at that time, the research of brain-computer interface has always been stagnated at the conceptual and theoretical stage. The scientific research team led by Professor Pfurtscheller in Austria is the pioneer and leader of BCI technology research. The team conducted a lot of research on the BCI system based on the EEG signal of motor imagery, and proposed the concepts of ERD and ERS for the first time (Pfurtscheller, 1992). In 2000, the team developed the Graz-BCI system based on the left and right hand motor imagery EEG signals and the motor sensory region mu and beta rhythm signals, and successfully realized the control of the cursor and the robotic arm (Pfurtscheller et al., 2002). Matsunaga et al. developed a wheelchair control system based on the EEG signal of motion imagination, which can perform basic movement and control operations (Tanaka et al., 2005). In 2008, the research team of the University of Tokyo in Japan used VR technology to conduct feedback training on subjects for the first time, which greatly improved the recognition rate of the subjects' motor imagination EEG signals (Fujisawa et al., 2008). The Birbaumer laboratory in Germany designed an SCP-based mind converter system. The user can use the EEG signal to control the input of characters, so as to achieve communication with the outside world (Birbaumer et al., 2003). In 2011, a research team from the Technical University of Berlin, Germany, developed a vehicle emergency brake assist system based on EEG signals. This system is 130 ms faster than manual operation and effectively shortens the braking distance (Haufe et al., 2011). In 2015, the Wadsworth Research Center in the United States and Karlsruhe Technical Research Institute in Germany jointly developed the "Brain-to-text" system (Herff et al., 2015). The system uses automatic speech recognition technology to convert human brain activity into corresponding text when speaking. This confirms the possibility of human-computer interaction based on natural speech-related cortical activity. In 2019, the team of Professor Bin He from Carnegie Mellon University in the United States and the University of Minnesota successfully developed the first "non-invasive mind control robotic arm" in history (Edelman et al., 2019). This is a non-invasively connected brain-computer interface system. Users can control the robotic arm to quickly track the randomly moving computer cursor through their mind and imagination. The system has achieved a nearly 60% improvement in learning performance, and the academic performance has also increased by more than five times compared with the previous one. In March 2020, Professor Edward Chang used Recursive Neural Network technology to learn neural features of EEG signals generated by epilepsy patients when they read aloud, and decode them into text sentences. The study achieved 97% accuracy (Makin et al., 2020).

Although China's sports imagination BCI technology started late, it has also made some progress. The BCI team led by Professor Gao Shangkai of Tsinghua University developed a BCI system in 2007 that uses left and right hand and foot motion imaginary EEG signals to control robotic dogs to complete football kicks (Wang et al., 2007). In 2010, the BCI team of South China University of Technology developed a hybrid BCI system that combined P300 and Mu/Beta rhythms to control the movement of a two-dimensional cursor (Li et al., 2010). In addition, the team used the same method to implement a wheelchair control BCI system in 2012 (Long et al., 2012). In 2015, Xu Baoguo and others from Southeast University developed an online robot control system based on MI-BCI

(Xu et al., 2015). On this platform, the feature extraction and classification of the EEG signal of hand movement imagination are realized, and the average classification accuracy rate reaches 91.5%. In 2016, Tang Zhichuan of Zhejiang University and others developed an exoskeleton upper limb rehabilitation robot based on brain control of motor imagination (Tang et al., 2016). The research achieved the highest online classification accuracy rate of 84.29 ± 2.11%. As an effective method to control the exoskeleton of the upper limbs, this research provides a non-invasive brain-controlled active upper limb rehabilitation strategy for clinical applications.

Many experts and scholars have achieved excellent results in the research of brain-computer interface systems and various analysis methods for EEG signals, which has also made a huge promotion for the scientific and effective analysis of EEG signals. Juneja et al. proposed an EEG classification method based on the ELM model, which is characterized by extracting the individual and mutual features of the data set (Juneja and Rana, 2019). Tan et al. proposed an improved ELM model for the classification of synchronous EEG (Tan et al., 2017). Zhang et al. used the differential entropy and attention model to classify EEG, so as to realize the automatic recognition of epilepsy (Zhang et al., 2020). However, the following problems still exist in the practicality of the brain-computer interface. First, the research on the classification of EEG signals for motor imaging mainly focuses on the two-classification problem, which is not enough to meet the actual needs. Second, the algorithm runs for a long time, and it is difficult to meet the user's requirements for timeliness. Third, the recognition rate needs to be further improved to improve the accuracy of issuing instructions. In response to these problems, this article intends to propose a motor imagery EEG classification method based on ELM algorithm. The main work of this paper is summarized as follows:

(1) The structure and operation process of the MI-BCI system are introduced. The system mainly includes three modules: preprocessing, feature extraction, and classification and recognition. The preprocessing steps in this article include data interception and band-pass filtering. Data interception refers to excluding the preparation and rest phases from the original EEG signal data segment, and only retains the motor imagination phase. Feature extraction includes lead selection, Local Characteristic-scale Decomposition (LCD) feature extraction, and Common Spatial Pattern (CSP) feature extraction. The classification recognition method is the ELM regression model, and the classification is performed by the regression method.
(2) A feature fusion method is used in the feature extraction module. The time domain features obtained based on the LCD feature extraction method and the spatial domain features obtained based on the CSP feature extraction method are combined in series to obtain a fusion feature.
(3) The classic ELM model is used for classification and recognition. The input weight matrix and hidden neuron bias matrix of ELM's hidden layer is randomly generated and obey any continuous probability distribution, so only the output weight matrix is required to be solved, and the input weight and hidden layer bias need not be adjusted iteratively. Therefore, the calculation amount and time complexity of the algorithm will be much smaller, and the training speed will be much faster. Recognition based on this model is more suitable for tasks that require high response time.

MI-BCI SYSTEM STRUCTURE

The MI-BCI system has different practical applications in different fields and scenarios, but a complete MI-BCI system is basically composed of five parts as shown in **Figure 1**.

(1) EEG signal acquisition. EEG signal acquisition is mainly responsible for EEG signal acquisition and storage. EEG signal acquisition methods can be divided into two types: non-implantable and implantable. In actual research, non-implantable acquisition methods are used in most cases, that is, the acquisition electrode is directly placed on the corresponding position of the subject's scalp to record the EEG signal. At the same time, the collected EEG signal is amplified and processed and A/D converted, and the processed result is stored.
(2) Pretreatment. The signal-to-noise ratio of the EEG signal is very low, and it is easily affected by external noise and other biological signals. It is very necessary to preprocess the EEG signal before analyzing and processing it, so as to reduce the artifact interference in the signal and improve the signal quality.
(3) Feature extraction. Feature extraction is one of the most important links in the MI-BCI system. Its purpose is to learn distinguishing features from the preprocessed EEG data, which is the feature that best reflects the true thinking activity of the brain. The commonly used methods include

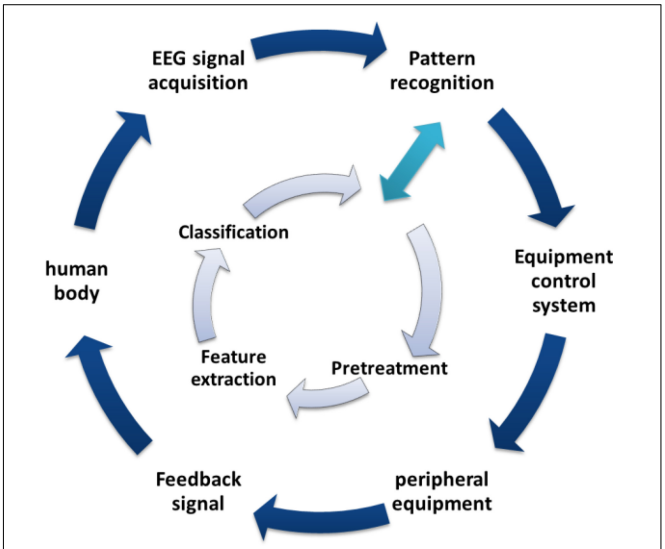


FIGURE 1 | MI-BCI system structure.

short-time Fourier transform, wavelet packet transform, common space mode algorithm, etc. (Cheng et al., 2019; Hu et al., 2019; Liu et al., 2020; Wang et al., 2020).

(4) Classification and identification. The classification and recognition of EEG signals is one of the key links that determines the performance of the BCI system. The role of this link is to integrate the extracted features and build a classification model. The classification model can establish a mapping relationship between the signal features and the subjects' real consciousness activities, so as to realize the feature classification of EEG signals. For EEG signals with different characteristics, commonly used classifiers include Fisher discriminant classifier, support vector machine, neural network algorithm, etc. (Rozza et al., 2012; Rubén et al., 2019; Dib et al., 2020; Ghonchi et al., 2020; Luo et al., 2020; Srirangan et al., 2020).

(5) Equipment control system. The function realized by the equipment control system is to convert the classification and recognition results of EEG signals into control commands for peripheral equipment. Common peripheral devices include character input systems, smart wheelchairs, robotic arms, etc.

ELM-BASED MOTOR IMAGE RECOGNITION METHOD

Identification Process

The identification process is shown in **Figure 2**. First, preprocess the collected motor imaging EEG signals. The preprocessing methods include data interception and band-pass filtering. Data interception removes the original EEG signal data segment from the preparation and rest phases, and only retains the motor imagination phase. Band-pass filtering retains the effective components of the EEG signal in the frequency range of 0.5–30 Hz, and removes irrelevant EMG and ECG components. Second, feature extraction. The time-frequency energy feature and spatial domain feature of EEG signal are extracted, and the features are fused. Third, the extracted features are input into the trained ELM model to obtain the classification result. Fourth, the device is controlled by the mapping relationship between the classification result and the control instruction.

Feature Extraction

It is necessary to select the lead before feature extraction. The pre-processed signals are calculated and sorted by the contribution rate, and the four leads with the highest information validity are distinguished from the remaining 18 leads. LCD (Yang et al., 2012) is used to decompose four lead EEG signals, 12 intrinsic scale components (ISC) are obtained, and the time-frequency energy feature vector V1 is obtained. To obtain the time-frequency energy of 12 ISC components, the expression of the time-frequency energy is as follows

$$E_{ij} = \sum_{t=0}^{N-1} \left| ISC_{ij}(t) \right|^2, i = 1, 2, 3, 4, j = 1, 2, 3 \quad (1)$$

$ISC_{ij}(t)$ refers to the first three layers of ISC components with length N obtained by decomposing each lead. Through the above processing, the time-frequency domain feature vector group $V_1=[H_{11},\ldots,H_{13},\ldots,H_{43}]$ is obtained. CSP (Dornhege et al., 2004) decomposes the remaining 18 leads EEG signals to obtain the spatial domain feature vector V2. A series fusion strategy is adopted to connect V1 and V2. The multiple features are normalized first, and then concatenated end to end. The mathematical expression is as follows:

$$\begin{cases} V_1 = \left[\frac{H_{11}}{|H_{11}|}, \ldots, \frac{H_{13}}{|H_{13}|}, \ldots, \frac{H_{43}}{|H_{43}|} \right] \\ V_2 = \left[\frac{f_1}{|f_1|}, \frac{f_2}{|f_2|}, \frac{f_3}{|f_3|}, \frac{f_4}{|f_4|} \right] \\ V = [V_1, V_2] \end{cases} \quad (2)$$

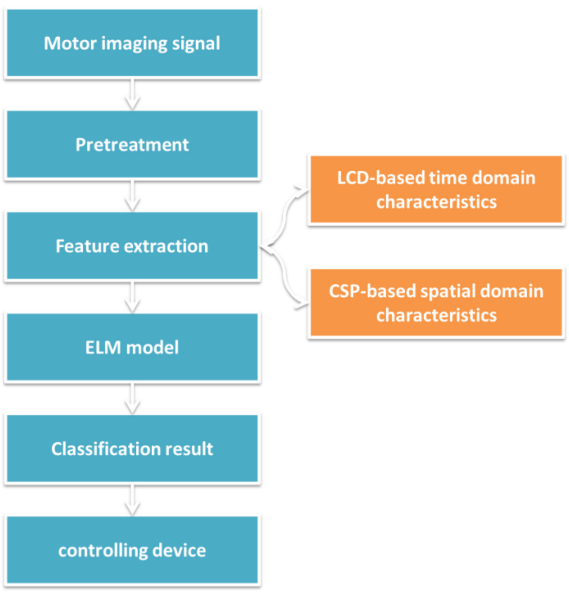


FIGURE 2 | Motion image recognition process based on ELM.

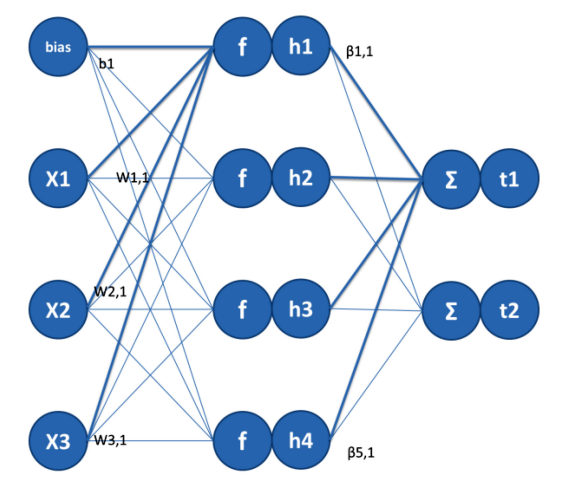


FIGURE 3 | Motion image recognition process based on ELM.

TABLE 1 | Data set introduction.

Data set	Subjects	Tasks	Channels	Rate	Experiments	Years
Data Sets IIb (Michael et al., 2012)	9	2 (left hand right hand)	3	250	400	2008
Data Sets IIIa (Schlögl et al., 2005)	3	4 (Left hand, right hand, tongue, feet)	60	250	240	2005

TABLE 2 | Classification accuracy of data sets IIb.

Subject	Feature extraction	SVM	Bayesian	TSK	ELM
1	LCD	0.7374	0.7288	0.7392	0.7408
	CSP	0.7286	0.7074	0.7111	0.7394
	LCD+CSP	0.7421	0.7320	0.7447	0.7532
2	LCD	0.6879	0.6724	0.6854	0.6858
	CSP	0.6580	0.6693	0.6622	0.6741
	LCD+CSP	0.7002	0.6798	0.6905	0.6998
3	LCD	0.7896	0.7669	0.7987	0.8008
	CSP	0.7744	0.7602	0.7880	0.7983
	LCD+CSP	0.7901	0.7754	0.8022	0.8112
4	LCD	0.7417	0.7491	0.7488	0.7504
	CSP	0.7289	0.7352	0.7306	0.7422
	LCD+CSP	0.7498	0.7594	0.7523	0.7610
5	LCD	0.7835	0.7713	0.7864	0.7915
	CSP	0.7749	0.7645	0.7800	0.7798
	LCD+CSP	0.7906	0.7802	0.7939	0.8003
6	LCD	0.7965	0.7843	0.7892	0.8002
	CSP	0.7893	0.7748	0.7816	0.7918
	LCD+CSP	0.7998	0.7841	0.7889	0.7988
7	LCD	0.8117	0.7993	0.7943	0.8016
	CSP	0.8023	0.7892	0.7868	0.7994
	LCD+CSP	0.8108	0.8034	0.7979	0.8180
8	LCD	0.7952	0.7867	0.8032	0.8128
	CSP	0.7794	0.7772	0.7804	0.7950
	LCD+CSP	0.7864	0.7893	0.8089	0.8236
Avg	LCD	0.7679	0.7574	0.7682	0.7730
	CSP	0.7545	0.7472	0.7526	0.7650
	LCD+CSP	0.7712	0.7630	0.7724	0.7832

TABLE 3 | Kappa coefficient of Data Sets IIb.

Feature extraction	SVM	Bayesian	TSK	ELM
LCD	0.5017	0.4841	0.5208	0.5619
CSP	0.4986	0.4749	0.5004	0.5572
LCD+CSP	0.5135	0.4981	0.5337	0.5897

ELM Classification Model

Single hidden layer feed-forward neural networks (SLFN) has been widely used because of its very strong nonlinear approximation ability. The latest research progress on SLFN learning is the Extreme Learning Machine (ELM) algorithm (Tamura and Tateishi, 1997; Huang, 2003; Huang et al., 2004, 2006). ELM is a single-layer feed forward network with fast training. There are only three layers in the network: The Input Layer, the Hidden Layer and the Output Layer. **Figure 3** shows the three-tier structure of the network.

In the Input Layer, each sample of each training set will have a corresponding weight and offset. There are two ways in ELM: one is to manually enter these weights and offsets, and the other is to automatically generate weights and offsets through ELM toolbox. Random generation is based on the size of the original data and the sigmoid neuron function. For a linear Output Layer, randomly generated weights can better reflect the performance advantages of ELM. ELM is a regression model, but it is also suitable for classification. If the different categories are separate and independent, create a target for each category separately. The target of the class that can be matched is set to 1, and the target of the unmatched class is set to 0. This encoding creates a unit-length vector for each category. This vector is orthogonal to all other classes of vectors. The distance between the target vectors of different categories is the same, so the independence of the categories can be maintained. The prediction category is assigned based on the target with the largest ELM output.

Literature (Huang et al., 2006) pointed out and proved that if the activation function $g(x)$ is infinitely differentiable, then, for a given arbitrarily small approximation error, the input layer weights and hidden layer thresholds arbitrarily select the number of hidden nodes $N' \leq N$, where N is the number of training samples. According to the above theory, when the activation function $g(x)$ is infinitely differentiable, all network parameters do not need to be adjusted. The input weight w_i and the hidden layer threshold b_i can be randomly assigned during training, and they are fixed during the training process. For the selected sample data, the hidden layer output matrix is fixed. Therefore, the parameter training of ELM is transformed into solving the following linear regression problem:

$$f(x) = \beta^T g \tag{3}$$

Here, β can be obtained by solving the least squares solution of the linear system $H\beta = T$. For the multiple-input single-output data set $D=\{(x_n, t_n) | x_n \in R^p, t_n \in R, n=1,2,...,N.\}$, the following optimization problems can be solved:

$$\min_{\beta} ||H\beta - T|| \tag{4}$$

where $H = \begin{bmatrix} g(w_1^T x_1 + b_1) & \cdots & g(w_{N'}^T x_1 + b_{N'}) \\ \cdots & \cdots & \cdots \\ g(w_1^T x_N + b_1) & \cdots & g(w_{N'}^T x_N + b_{N'}) \end{bmatrix}_{N \times N'} = \begin{bmatrix} g^T(x_1) \\ \cdots \\ g^T(x_N) \end{bmatrix}_{N \times N'}$ is the hidden output matrix. $\beta = [\beta_1, \beta_2, ..., \beta_{N'}]^T$ is the output node, $T = [t_1, t_2, \cdots, t_N]^T$ is the hidden layer connection weight vector. The least square solution of Eq. (3) is

$$\hat{\beta} = H^+ T \tag{5}$$

where H^+ is the Moore-Penrose generalized inverse of the hidden output matrix H. The optimal solution A has the following

important characteristics (1) The smallest training error can be obtained through this solution; (2) The weight vector of the smallest normal form is obtained; (3) The least squares solution of the paradigm is unique, so the algorithm will not produce a local optimal solution.

EXPERIMENTAL PROCESS AND RESULTS ANALYSIS

Experimental Data Set

The data set used in this article is the BCI competition data set provided by Graz University. The description of the experimental data is shown in **Table 1**.

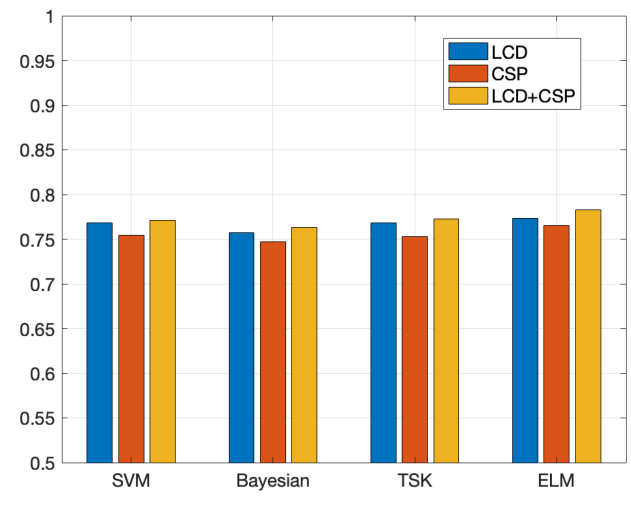


FIGURE 4 | Classification accuracy obtained by each classification model on the Data Sets IIb.

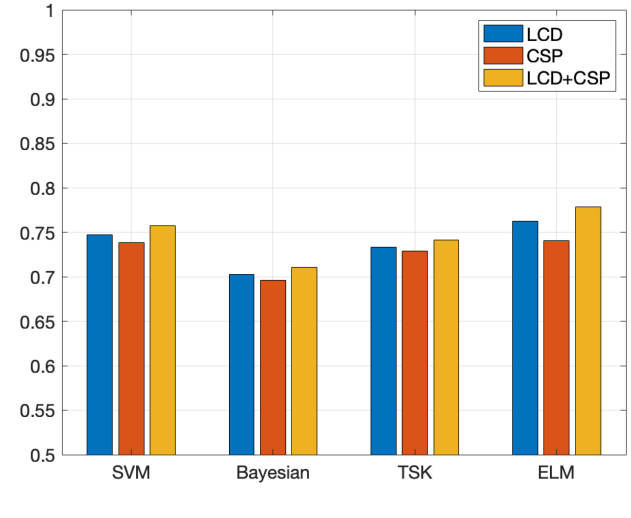


FIGURE 5 | Kappa coefficient obtained by each classification model on the Data Sets IIb.

Experimental Setup

To verify the effectiveness of the method used in this paper, the selected contrast classifiers include support vector machine SVM (Shahyad et al., 2019), Bayesian classifier (Bayesian) (Zhang et al., 2016), Takagi–Sugeno–Kang (TSK) (Gu et al., 2017). To verify the superiority of the fusion features, the individual time-frequency features and spatial domain features were compared with the fusion features of the two in the experiment process. This paper uses the classification accuracy rate acc and the kappa coefficient to measure the classification accuracy to evaluate the classification performance of the model. The larger the kappa coefficient, the better the algorithm performance. The mathematical expression of the kappa coefficient is as follows:

$$Kappa = \frac{acc - 1/N}{1 - 1/N} \tag{6}$$

where *acc* represents the classification accuracy rate, and N represents the number of categories. The value range of kappa is [0,1]. The larger the value, the stronger the consistency and the higher the classification accuracy. During the experiment, the data set was divided into training set and test set. The training set is used to build the classifier model, and the test set is used to evaluate the accuracy of the model for predictive classification of unknown samples. To scientifically test the classification accuracy rate and avoid misleading the experimental results caused by a single classification result, this article adopts fivefold cross validation. Each time, the data set is divided into five parts, and four of them are used as training data to obtain the model completely randomly. The remaining one is classified as test data. The trained classifier model is used to classify the test set. From the classification results, the accuracy rate acc_i (i = 1, 2,..., 5) and the Kappa coefficient K_i (i = 1, 2,..., 5) can be obtained. The above

TABLE 4 | Classification accuracy of data sets IIIa.

Subject	Feature extraction	SVM	Bayesian	TSK	ELM
1	LCD	0.8257	0.8031	0.8209	0.8286
	CSP	0.8135	0.8059	0.8187	0.8145
	LCD+CSP	0.8294	0.8202	0.8231	0.8466
2	LCD	0.8121	0.7985	0.8073	0.8252
	CSP	0.8093	0.7917	0.8028	0.8565
	LCD+CSP	0.8289	0.8123	0.8162	0.8334
3	LCD	0.7982	0.7645	0.8014	0.8105
	CSP	0.7819	0.7492	0.7906	0.8067
	LCD+CSP	0.7998	0.7685	0.8182	0.8242
Avg	LCD	0.8120	0.7887	0.8099	0.8214
	CSP	0.8016	0.7823	0.8040	0.8259
	LCD+CSP	0.8194	0.8003	0.8192	0.8347

TABLE 5 | Kappa coefficient of data sets IIIa.

Feature extraction	SVM	Bayesian	TSK	ELM
LCD	0.7472	0.7022	0.7331	0.7623
CSP	0.7380	0.6958	0.7290	0.7405
LCD+CSP	0.7572	0.7105	0.7412	0.7788

process is repeated five times, and the average value of five times is used as the final accuracy of the algorithm. The final calculation formula of the experimental results is as follows, where F is 5.

$$\bar{A} = \frac{1}{F}\sum_{i=1}^{F} acc_i \tag{7}$$

$$\bar{K} = \frac{1}{F}\sum_{i=1}^{F} K_i \tag{8}$$

Discussion of Experimental Results

(1) Data Sets IIb

In order to demonstrate the effectiveness of the method in this paper, this paper compares feature extraction and classifier performance. The experimental results on the Data Sets IIb data set are shown in **Tables 2**, **3**. **Figures 4**, **5** show the classification accuracy and Kappa coefficient obtained by each classification model on the Data Sets IIb data set. The classification on this data set is a binary classification problem.

The experimental results on Data Sets II give two feedback points. (1) The three features are the time domain features based on the LCD feature extraction method, the spatial domain features based on the CSP feature extraction method, and the fusion feature obtained by combining the time domain feature and the spatial domain feature. The classification results obtained based on the above three features show that the classification effect obtained by fusing the features is the best. Different feature extraction methods perform basically the same on the four classification algorithms. For each classifier, the classification effect based on fusion features is the best. This shows that the information of fusion features is richer and more comprehensive, which is more conducive to classification tasks. In most classification algorithms, the classification effect based on the LCD feature extraction method is better than the classification effect based on the CSP feature extraction method. This shows that time domain features have higher information value than spatial domain features. (2) Comparing the four classification algorithms, the ELM algorithm used in this article has the best performance, followed by TSK, third is SVM, and the worst is Bayesian. For other types of classification algorithms, ELM can initialize the input weights and biases randomly to get the corresponding output weights. This algorithm can be faster than traditional learning algorithms under the premise of ensuring learning accuracy. This is the reason why this algorithm is chosen as the classification model in this article.

(2) Data Sets IIIa

The classification results on the Data Sets IIIa data set are shown in **Tables 4**, **5**. **Figures 6**, **7** show the classification accuracy and Kappa coefficient obtained by each classification model on the Data Sets IIIa data set. The classification of this data set belongs to four classifications.

The data set on Data Sets IIIa is classified into four categories, and the experimental results obtained are somewhat different from the experimental results on Data Sets II. By analyzing the experimental data in **Tables 4**, **5**, the following conclusions are obtained. (1) The fusion feature extraction method among the three feature extraction methods has the best classification effect. On this data set, the classification effect based on the LCD feature extraction method is better than the classification effect based on the CSP feature extraction method, but the gap in classification accuracy is not very large. (2) The classification effect of the ELM algorithm is still the best, followed by SVM, TSK third, and Bayesian the worst. Compared with the experimental results obtained from the Data Sets II data set, the position of the best classification effect of ELM has not changed, but the classification effect of SVM classifier surpasses TSK. This fully demonstrates the feasibility and superiority of selecting ELM as the classification algorithm in this paper. Compared with other classification algorithms, the classification performance of the ELM algorithm is more stable and the advantages are more obvious.

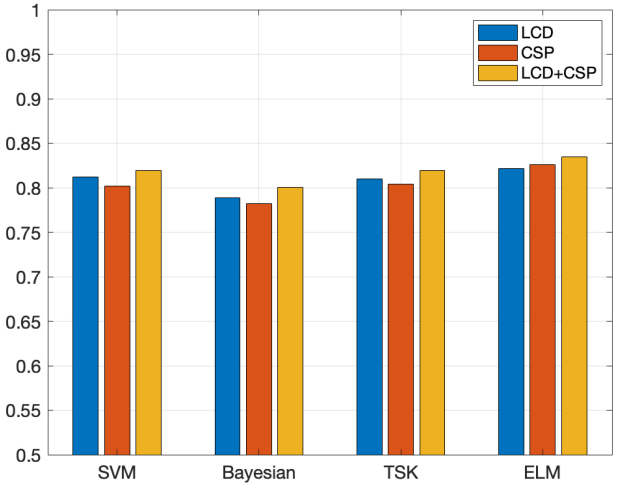


FIGURE 6 | Classification accuracy obtained by each classification model on the Data Sets IIIa.

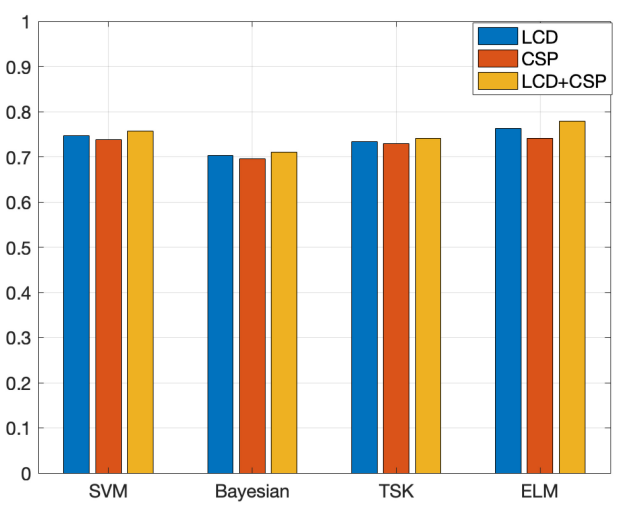


FIGURE 7 | Kappa coefficient obtained by each classification model on the Data Sets IIIa.

CONCLUSION

A novel smart motor imagery intention human-computer interaction model using extreme learning machine and EEG signals is proposed. Aiming at the problems of low recognition accuracy, large individual differences, and feature redundancy in the current multi-classification problem of motor imagery EEG, this paper deeply studies the signal recognition methods of different limb motor imagery to improve the application of BCI system. The main work of this paper is summarized as follows: (1) A multi-feature fusion extraction method, namely the feature extraction method of LCD+CSP, is used. This method can extract the time-frequency domain and spatial domain features of the data set. The experimental results show that the classification result obtained by a single feature is not as good as the classification result obtained by the fusion feature. (2) The selected ELM algorithm is a mature and widely used classification algorithm with relatively high classification accuracy and fast running speed. ELM can initialize the input weights and biases randomly to get the corresponding output weights. This algorithm can be faster than traditional learning algorithms under the premise of ensuring learning accuracy. The experimental results verify that the ELM algorithm has certain classification advantages. The motion image intention recognition effect based on the method in this paper is close to the champion algorithm of BCI competition. Therefore, the human-computer interaction model designed based on the recognition results of this method is feasible to help patients recover. However, the data set used in this article is relatively simple, and they are all data sets in the BCI competition. In the follow-up, this research plans to use richer data sets for experimental research to verify the universality of the algorithm in this paper. The key to the effectiveness of the ELM model is whether it is necessary to extract features from the data. If feature extraction is meaningless, ELM can come in handy. If it is for the original data set, and feature extraction is very important, the classification effect of ELM is not ideal. In addition, ELM sacrifices too many meaningful patterns in data for speed. Therefore, how to improve ELM to determine the balance between classification accuracy and speed is a future research work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.bbci.de/competition/iv/ and http://www.bbci.de/competition/iii/.

AUTHOR CONTRIBUTIONS

YG developed the theoretical framework and model in this work and drafted the manuscript. YG and LH implemented the algorithm and performed experiments and result analysis. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61772241 and 61873321 and in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant No. XYDXX-127.

REFERENCES

Birbaumer, N., Hinterberger, T., and Kubler, A. (2003). The thought-translation device (ttd):neurobehavioral mechanisms and clinical outcome. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 120–123. doi: 10.1109/tnsre.2003.814439

Cheng, L., Li, D., Li, X., and Yu, S. (2019). The optimal wavelet basis function selection in feature extraction of motor imagery electroencephalogram based on wavelet packet transformation. *IEEE Access* 7, 174465–174481. doi: 10.1109/access.2019.2953972

Dib, N., Radhwane, B., and Fethi, B. R. (2020). Epileptic seizure recognition using EEG wavelet decomposition based on nonlinear and statistical features with support vector machine classification. *Biomed. Eng.* 65, 133–148. doi: 10.1515/bmt-2018-0246

Dornhege, G., Blankertz, B., Curio, G., and Mller, K. R. (2004). Increase information transfer rates in BCI by CSP extension to multi-class. *Adv. Neual Inf. Process. Syst.* 2004, 733–740.

Edelman, B. J., Meng, J., Suma, D., Zurn, C., Nagarajan, E., Baxter, B. S., et al. (2019). Noninvasive neuroimaging enhances continuous neural tracking for robotic device control. *Sci. Robot.* 4:6844. doi: 10.1126/scirobotics.aaw6844

Fujisawa, J., Touyama, H., and Hirose, M. (2008). EEG-based navigation of immersing virtual environment using common spatial patterns. *IEEE Virtual Real. Conf.* 2008, 251–252.

Ghonchi, H., Fateh, M., Abolghasemi, V., Ferdowsi, S., and Rezvani, M. (2020). Deep recurrent-convolutional neural network for classification of simultaneous EEG-fNIRS signals. *IET Signal Process.* 14, 142–153. doi: 10.1049/iet-spr.2019.0297

Gu, X., Chung, F. L., and Wang, S. (2017). Bayesian Takagi–Sugeno–Kang Fuzzy Classifier. *IEEE Trans. Fuzzy Syst.* 99, 1655–1671. doi: 10.1109/tfuzz.2016.2617377

Hammer, E. M., Halder, S., Kleih, S. C., and Kübler, A. (2018). Psychological predictors of visual and auditory P300 brain-computer interface performance. *Front. Neurosci.* 12:307. doi: 10.3389/fnins.2018.00307

Haufe, S., Treder, M. S., and Gugler, M. F. (2011). EEG potentials predict upcoming emergency brakings during simulated driving. *J. Neural Eng.* 8, 1741–1746. doi: 10.1088/1741-2560/8/5/056001

Herff, C., Heger, D., and Pesters, A. D. (2015). Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front. Neurosci.* 9:217. doi: 10.3389/fnins.2015.00217

Hu, J., Luo, Y., Zhang, Q., Liu, X., and Qu, X. (2019). Micro-motion feature extraction of space targets based on sinusoidal frequency modulation Fourier transform. *J. Eng.* 2019, 8076–8079. doi: 10.1049/joe.2019.0650

Huang, G. B. (2003). Learning capability and storage capacity of two-hidden-layer feed-forward networks. *IEEE Trans. Neural Netw.* 14, 274–281. doi: 10.1109/tnn.2003.809401

Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2004). "Extreme learning machine: a new learning scheme of feed-forward neural networks," in *Proceedings of IEEE International Joint Conference on Neural Networks*, (Piscalaway, NJ: IEEE Press), 985–990. doi: 10.1007/s10916-017-0867-4

Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing* 70, 489–501. doi: 10.1016/j.neucom.2005.12.126

Juneja, K., and Rana, C. (2019). Individual and mutual feature processed ELM model for EEG signal based brain activity classification. *Wirel. Pers. Commun.* 108, 659–681. doi: 10.1007/s11277-019-06423-w

Li, Y., Long, J., and Yu, T. (2010). An EEG-based BCI system for 2-D cursor control by combining Mu/Beta rhythm and P300 potential. *IEEE Trans. Biomed. Eng.* 57, 2495–2505. doi: 10.1109/tbme.2010.2055564

Lisi, G., Noda, T., and Morimoto, J. (2014). Decoding the ERD/ERS: influence of afferent input induced by a leg assistive robot. *Front. Syst. Neurosci.* 8:85. doi: 10.3389/fnsys.2014.00085 doi: 10.3389/fnsys.2014.00085

Liu, C., Shi, D., and Wang, Z. (2020). Acoustic emission damage feature extraction of refractories based on wavelet packet transform and independent component analysis. *Naihuo Cailiao Refractories* 54, 185–190.

Long, J., Li, Y., and Wang, H. (2012). A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 720–729. doi: 10.1109/tnsre.2012.2197221

Luo, Y., Fu, Q., Xie, J., Qin, Y., Wu, G., Liu, J., et al. (2020). EEG-based emotion classification using spiking neural networks. *IEEE Access* 8, 46007–46016. doi: 10.1109/access.2020.2978163

Makin, J. G., Moses, D. A., and Chang, E. F. (2020). Machine translation of cortical activity to text with an encoder–decoder framework. *Nat. Neurosci.* 23, 575–582. doi: 10.1038/s41593-020-0608-8

Michael, T., Müller, K. R., and Ad, A. (2012). Review of the BCI Competition IV. *Front. Neurosci.* 6:55. doi: 10.3389/fnins.2012.00055

Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. *Electroencephalogr. Clin. Neurophysiol.* 83, 62–69. doi: 10.1016/0013-4694(92)90133-3

Pfurtscheller, G., Neuper, C., and Guger, C. (2002). Current trends in Graz brain-computer interface (BCI) research. *IEEE Trans. Rehabil. Eng.* 8, 216–219. doi: 10.1109/86.847821

Qiu, Z., Allison, B. Z., and Jin, J. (2017). Optimized motor imagery paradigm based on imagining Chinese characters writing movement. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1009–1017. doi: 10.1109/tnsre.2017.2655542

Rozza, A., Lombardi, G., and Casiraghi, E. (2012). Novel Fisher discriminant classifiers. *Pattern Recogn.* 45, 3725–3737. doi: 10.1016/j.patcog.2012.03.021

Rubén, S. S., Manuel, G. M., Luis Fernando, D. E., and José, M. P. (2019). Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. *Comput. Biol. Med.* 109, 148–158. doi: 10.1016/j.compbiomed.2019.04.031

Schlögl, A., Lee, F., and Bischof, H. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *journal of neural engineering* 2, L14. doi: 10.1088/1741-2560/2/4/L02

Shahyad, S., Hatef, B., Khosrowabadi, R., Mohammadi, A., and Lotfan, S. (2019). Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures. *Biocybern. Biomed. Eng.* 39, 199–213. doi: 10.1016/j.bbe.2018.10.008

Srirangan, M., Rajesh, K. T., and Ram, B. P. (2020). Time-frequency domain deep convolutional neural network for the classification of focal and non-focal EEG signals. *IEEE Sensors J.* 20, 3078–3086. doi: 10.1109/jsen.2019.2956072

Tamura, S., and Tateishi, M. (1997). Capability of four-layered neural network: four layers versus three. *IEEE Trans. Neural Netw.* 8, 251–255. doi: 10.1109/72.557662

Tan, P., Tan, G. Z., Cai, Z. X., Sa, W. P., and Zou, Y. Q. (2017). Using ELM-based weighted probabilistic model in the classification of synchronous EEG BCI. *Med. Biol. Eng. Comput.* 55, 33–43. doi: 10.1007/s11517-016-1493-x

Tanaka, K., Matsunaga, K., and Wang, H. O. (2005). Electroencephalogram-based control of an electric wheelchair. *IEEE Trans. Robot.* 21, 762–766. doi: 10.1109/tro.2004.842350

Tang, Z., Sun, S., Zhang, S., Chen, Y., and Li, C. (2016). A brain-machine interface based on ERD/ERS for an upper-limb exoskeleton control. *Sensors* 16:2050. doi: 10.3390/s16122050

Van Dokkum, L. E. H., Ward, T., and Laffont, I. (2015). Brain computer interfaces for neurorehabilitation-its current status as a rehabilitation strategy post-stroke. *Ann. Phys. Rehabil. Med.* 58, 3–8. doi: 10.1016/j.rehab.2014.09.016

Vidal, J. J. (1973). Toward direct brain-computer communication. *Annu. Rev. Biophys. Bioeng.* 2, 157–180. doi: 10.1146/annurev.bb.02.060173.001105

Wang, W., Sun, L., Wang, B., and Yu, M. (2020). Partial discharge feature extraction based on synchrosqueezed windowed Fourier transform and multi-scale dispersion entropy. *Meas. Control* 53, 1078–1087. doi: 10.1177/0020294020932346

Wang, Y., Hong, B., and Gao, X. (2007). "Implementation of a brain-computer interface based on three states of motor imagery," in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, (Lyon: IEEE), 5059–5062.

Wolpaw, J. R., Birbaumer, N., and Heetderks, W. J. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/tre.2000.847807

Xu, B., Song, A., Zhao, G., Xu, G., Pan, L., Yang, R., et al. (2015). Design and evaluation of a motor imagery electroencephalogram-controlled robot system. *Adv. Mech. Eng.* 7, 8132–8140.

Yang, Y., Zeng, M., and Cheng, J. S. (2012). Research on local characteristic-scale decomposition and its capacities. *J. Vib. Eng.* 25, 602–609.

Zhang, Y., Zhou, G. X., Jin, J., Zhao, Q. B., Wang, X. Y., and Cichocki, A. (2016). Sparse Bayesian Classification of EEG for brain–computer interface. *IEEE Trans. Neural Netw. Learn. Syst.* 27, 2256–2267. doi: 10.1109/tnnls.2015.2476656

Zhang, J., Wei, Z., Zou, J. , and Fu, H. (2020). Automatic epileptic eeg classification based on differential entropy and attention model. *Eng. Appl. Artif. Intell.* 96:103975. doi: 10.1016/j.engappai.2020.103975

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 07 May 2021
doi: 10.3389/fnins.2021.634063

Advanced Analysis of Diffusion Tensor Imaging Along With Machine Learning Provides New Sensitive Measures of Tissue Pathology and Intra-Lesion Activity in Multiple Sclerosis

*Olayinka Oladosu[1,2], Wei-Qiao Liu[2,3], Bruce G. Pike[2,4], Marcus Koch[2,3], Luanne M. Metz[2,3] and Yunyan Zhang[2,3,4]**

[1] Department of Neuroscience, Faculty of Graduate Studies, University of Calgary, Calgary, AB, Canada, [2] Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada, [3] Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada, [4] Department of Radiology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:
Yu-Dong Zhang, University of Leicester, United Kingdom

Reviewed by:
Mojgan Hodaie, University of Toronto, Canada
Karl Helmer, Massachusetts General Hospital and Harvard Medical School, United States

***Correspondence:**
Yunyan Zhang
yunyzhan@ucalgary.ca

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *26 November 2020*
Accepted: *15 April 2021*
Published: *07 May 2021*

Citation:
Oladosu O, Liu W-Q, Pike BG, Koch M, Metz LM and Zhang Y (2021) Advanced Analysis of Diffusion Tensor Imaging Along With Machine Learning Provides New Sensitive Measures of Tissue Pathology and Intra-Lesion Activity in Multiple Sclerosis. Front. Neurosci. 15:634063. doi: 10.3389/fnins.2021.634063

Tissue pathology in multiple sclerosis (MS) is highly complex, requiring multi-dimensional analysis. In this study, our goal was to test the feasibility of obtaining high angular resolution diffusion imaging (HARDI) metrics through single-shell modeling of diffusion tensor imaging (DTI) data, and investigate how advanced measures from single-shell HARDI and DTI tractography perform relative to classical DTI metrics in assessing MS pathology. We examined 52 relapsing-remitting MS patients who had 3T anatomical brain MRI and DTI. Single-shell HARDI modeling yielded 5 sub-voxel-based metrics, totalling 11 diffusion measures including 4 DTI and 2 tractography metrics. Based on machine learning of 3-dimensional regions of interest, we evaluated the importance of the measures through several tissue classification tasks. These included two within-subject comparisons: lesion versus normal appearing white matter (NAWM); and lesion core versus shell. Further, by stratifying patients as having high (above 75%^{ile}) and low (below 25%^{ile}) number of MS lesions, we also performed 2 classifications between subjects for lesions and NAWM respectively. Results showed that in lesion-NAWM analysis, HARDI orientation distribution function (ODF) energy, DTI fractional anisotropy (FA), and HARDI orientation dispersion index were the top three metrics, which together achieved 65.2% accuracy and 0.71 area under the receiver operating characteristic curve (AUROC). In core-shell analysis, DTI mean diffusivity (MD), radial diffusivity, and FA were the top three metrics, and MD dominated the classification, which achieved 59.3% accuracy and 0.59 AUROC alone. Between patients, FA was the leading feature in lesion comparisons, while ODF energy was the best in NAWM separation. Collectively,

single-shell modeling of common diffusion data can provide robust orientation measures of lesion and NAWM pathology, and DTI metrics are most sensitive to intra-lesion abnormality. Combined analysis of both advanced and classical diffusion measures may be critical for improved understanding of MS pathology.

Keywords: single-shell high angular resolution diffusion imaging, diffusion tensor imaging, tractography, support vector machine, lesions, intra-lesion pathology

INTRODUCTION

Multiple sclerosis (MS) is a severe central nervous system disease impacting > 2.8 million people worldwide (Msif Tmsif., 2020). Focal lesions are the hallmark of MS pathology, characterized by several changes, including demyelination, axonal injury, and inflammation (Reich et al., 2018). In addition, MS pathology alters tissue integrity not only in the normal-appearing tissue, but also within the territory of established lesions, leading to heterogeneous damage as seen in chronic active lesions. Ongoing tissue damage in the latter is believed to play a major role in the relentless progression of patient disability in MS (Reich et al., 2018). However, robust measurements of the complex types of MS pathology are still missing. Diffusion magnetic resonance imaging (dMRI) serves as a promising tool for *in vivo* assessment of tissue microstructure (Inglese and Bester, 2010). Moreover, with advances in dMRI techniques, multi-dimensional analysis of tissue properties becomes possible, including localized analysis of lesion activities concerning specific white matter tracts (Inglese and Bester, 2010).

Diffusion tensor imaging (DTI) is a common dMRI method that reflects the orientational dependence of diffusion associated with individual image voxels (Tournier, 2019). Diffusion tensor imaging uses a tensor model, which defines diffusion anisotropy according to three perpendicular diffusion axes originating from a voxel (Tournier, 2019). Several studies have shown that the mean diffusivity (MD) and fractional anisotropy (FA) of DTI detect demyelination and axonal loss in MS, which differentiates lesions from the normal-appearing white matter (NAWM) (Schmierer et al., 2007). Within the lesion context, there is evidence showing increased MD in the lesion core versus perilesional white matter (Klistorner et al., 2018). However, there are no systemic studies of core versus periphery of lesions in MS, and use of a single tensor model in DTI, which does not provide compartmental information, limits its ability to detect specific processes of neuronal pathology. High angular resolution diffusion imaging (HARDI) tackles this problem by using advanced models which enable intra-voxel analysis through increased angular sampling of diffusion signals acquired typically with different diffusion weightings (multi-shell HARDI) (Descoteaux, 2015). A range of multi-shell HARDI models exist, divided mainly by the approach used to model the properties of nerve fibers, such as the cylinder model used in ActiveAx (Alexander et al., 2010), and the stick model in neurite orientation dispersion and density index (NODDI) (Zhang et al., 2012). But acquiring multi-shell HARDI data is not always practical due to time constraints, particularly in a clinical setting.

An alternative approach to multi-shell HARDI is modeling densely sampled data acquired using one diffusion weighting, namely, single-shell HARDI (ssHARDI) (Taquet et al., 2013). In general, dMRI acquired with gradient sampling orientations ≥ 45 is sufficient for this approach (Descoteaux, 2015). Current investigations of ssHARDI mainly focus on its neurite orientation characterization potential (Taquet et al., 2013). While some work has explored the possibility of generating additional compartmental metrics (Grussu et al., 2014; Magnollay et al., 2014), the ability of ssHARDI for advanced intra-voxel analysis of tissue properties remains unclear, particularly based on typical DTI data.

In addition to advances in diffusion modelling, there have also been considerable improvements in tractography investigations based on dMRI (Tournier, 2019). Traditionally, individual white matter tracts (e.g., corticospinal tract) traced by streamlines of tractography form the mainstay for localized analysis of diffusion metrics. However, the local orientation information represented directly by the streamlines may be also critical indicators of tissue properties, including streamline counts and streamline termination frequency, providing a new dimension of microstructural measurements (Ukmar et al., 2012).

The purpose of this study was to test the feasibility of ssHARDI using clinically available dMRI, and investigate how advanced metrics from ssHARDI and DTI tractography compare to traditional DTI measures in assessing MS pathology. The investigations used a machine learning technique, support vector machine (SVM), to evaluate feature importance through several tissue classification tasks. Specifically, based on 3D regions of interest (ROIs), there were two classifications within individual patients: (1) lesion versus the corresponding contralateral NAWM, to obtain a basic understanding of the sensitivity of the features to MS pathology; and (2) lesion core versus shell, to evaluate the detectability of intra-lesion pathology by the features. Between patients, there were also two analyses: lesion comparisons, between patients having high (top 25%) and low (bottom 25%) lesion counts as a surrogate for disease activity; and NAWM comparisons, between the same subject groups.

MATERIALS AND METHODS

Sample

This study evaluated brain magnetic resonance imaging (MRI) scans from a convenience sample of 52 participants with relapsing-remitting multiple sclerosis (RRMS) who were screened for participation in a clinical trial of domperidone as a potential myelin repair agent (ClinicalTrials.gov Identifier:

NCT02493049). Participants required at least one gadolinium-enhancing lesion on a screening brain MRI to be eligible for randomization to treatment in the clinical trial. In the present study, we utilized the screening brain MRI scans of participants who were not eligible to continue in the trial as they did not contain any enhancing lesions. In addition, all MRI pulse sequences used in this study were conducted before gadolinium use, and therefore no contrast interference. Written informed consent was obtained from all participants following study approval by the Conjoint Health Research Ethics Board of the University of Calgary.

Imaging Protocol

All participants had brain anatomical and diffusion MRI undertaken at a 3T scanner (GE Healthcare, Discovery MR750, Milwaukee, United States). Anatomical MRI included T2-weighted and FLAIR images with repetition time (TR) = 6000/7000, echo time (TE) = 100/126 ms, matrix = 512 × 512, and slice thickness = 3 mm without gap; and T1-weighted images with TR/TE = 8.2/3.2 ms, matrix = 256 × 256, field of view (FOV) = 25 × 25 cm, and slice thickness = 1 mm. The dMRI acquisition used an echo planar sequence where b = 1000, 3 b0 volumes, and 45 directions; TR/TE = 8000/61 ms; matrix = 120 × 120, with 2 mm (Inglese and Bester, 2010) isotropic voxels; and FOV = 24 × 24 cm.

3D ROI Development

To improve the analysis of contextual information, we derived 3D ROIs for all tissue regions, done initially using anatomical MRI (**Figure 1**).

Lesion ROIs

Using the FSL library (Oxford, United Kingdom), all MRI volumes were skull-extracted, and the T2 and FLAIR images were then rigid-body co-registered to T1 images to optimize quality and alignment (Jenkinson and Smith, 2001; Jenkinson et al., 2002; Jenkinson et al., 2012). Lesion segmentation focused on brain white matter, using an automatic toolbox (LST, v3.0.0, SPM12) (Schmidt et al., 2012). Initially, the LST generated a whole-brain lesion map per subject based on co-registered FLAIR and T1 volumes. The lesion map then underwent manual correction using ImageJ (NIH, v1.52j) by referencing the co-registered FLAIR and T2 images. Any area that contained a confluent lesion, showing signal inhomogeneity but with pixels staying connected, was considered a single ROI. Subsequently, lesion ROIs were colocalized across slices using the 'cluster' command in FSL with 26-connectivity to obtain 3D ROIs, which were further sorted by ROI size.

Contralateral NAWM ROIs

The NAWM ROIs were essentially the mirror image of the lesion ROIs established above. To ensure validity, we applied several quality-control steps. The first step was lesion-filling (Battaglini et al., 2012) in the referencing T1 MRI, followed by creation of a left-right flipped mirror image of the volume. The accuracy of volume flipping was ensured through a cross-correlation-based nonlinear co-registration process between the reference and mirror T1 volumes, using the "SyN" option in the ANTs software (Avants et al., 2008), from which a left-right transformation was obtained. Applying the transformation to the generated lesion masks made the latter geometrically matched to the mirror volume of the reference T1. After left-right flipping, the transformed lesion masks became the "raw" contralateral NAWM ROIs. The next step was refinement of the NAWM ROIs, including eliminating areas overlapping with any lesion region. This step ensured that each 3D NAWM ROI corresponded to each unique 3D lesion ROI with no contamination by tissues of the other type (see **Figure 1**).

Lesion Core and Shell ROIs

The core-shell segmentation focused on lesion ROIs that were large enough to encapsulate a 3 × 3 × 3 cube of voxels. Specifically, defining a lesion core ROI used a 3D volumetric erosion process applied to an eligible lesion mask, which allowed to retain only the central voxels not in contact with any non-lesion background area. Then, subtracting the core from the full lesion ROI produced the single-voxel-thick shell ROI for each lesion.

dMRI Analysis

Pre-processing and DTI Analysis

The dMRI data first underwent eddy current correction (Andersson and Sotiropoulos, 2016). To limit variations, we averaged the 3 b0 volumes and then co-registered the mean volume to the T1 structural space. The resulting mean b0 volume acted as a reference in transforming all calculated diffusion maps to the same T1 space prior to quantitative evaluations. DTI calculation used the FDT procedure in FSL, which provided 4 classical outcomes: MD, FA, axial diffusivity (AD), and radial diffusivity (RD).

ssHARDI Analysis

Given the relatively high angular nature of our dMRI acquisitions, we also explored ssHARDI modeling. In particular, to improve computing efficiency, we used a new modeling method: Accelerated Microstructure Imaging via Convex Optimization (AMICO), which generated equivalent measures to ActiveAx and NODDI (Daducci et al., 2015). Our AMICO outcomes included axonal diameter, axonal density, and intracellular volume fraction (ICVF) from ActiveAx, and orientation dispersion index (ODI) from NODDI (Alexander et al., 2010; Zhang et al., 2012). In addition, to further probe intravoxel orientation information beyond ODI, we calculated the orientation distribution function (ODF) of diffusion (Descoteaux et al., 2007). This in turn enabled us to generate a new parameter termed ODF energy (**Figure 2**), which referred to the energy of diffusion oriented at individual directions. The energy was calculated as log (p^2), where p represented a collection of probabilities of diffusion magnitude observed at all possible directions. The probabilities were obtained by fitting the diffusion magnitude values from each direction to a normal distribution.

In addition, to test the feasibility of using AMICO to evaluate ssHARDI, we performed an additional experiment to

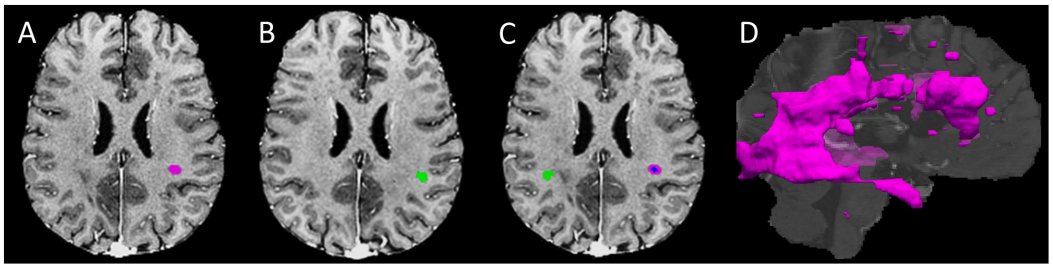


FIGURE 1 | Regions of interest development. **(A)** An example lesion region of interest (ROI) in the T1- FLAIR MRI space (purple mask); **(B)** A mirror-image of the T1-weighted MRI for developing the contralateral normal appearing white matter (NAWM) ROI (green mask) of the lesion shown in **panel A**; **(C)** the finalized contralateral NAWM ROI of the lesion produced using (**panel A,B**), along with converse co-registration between the corresponding image volumes; the blue and purple areas within the lesion represents the core and shell ROIs, respectively; and **(D)** 3D whole brain lesion masks produced by grouping adjacent ROI voxels within and between slices.

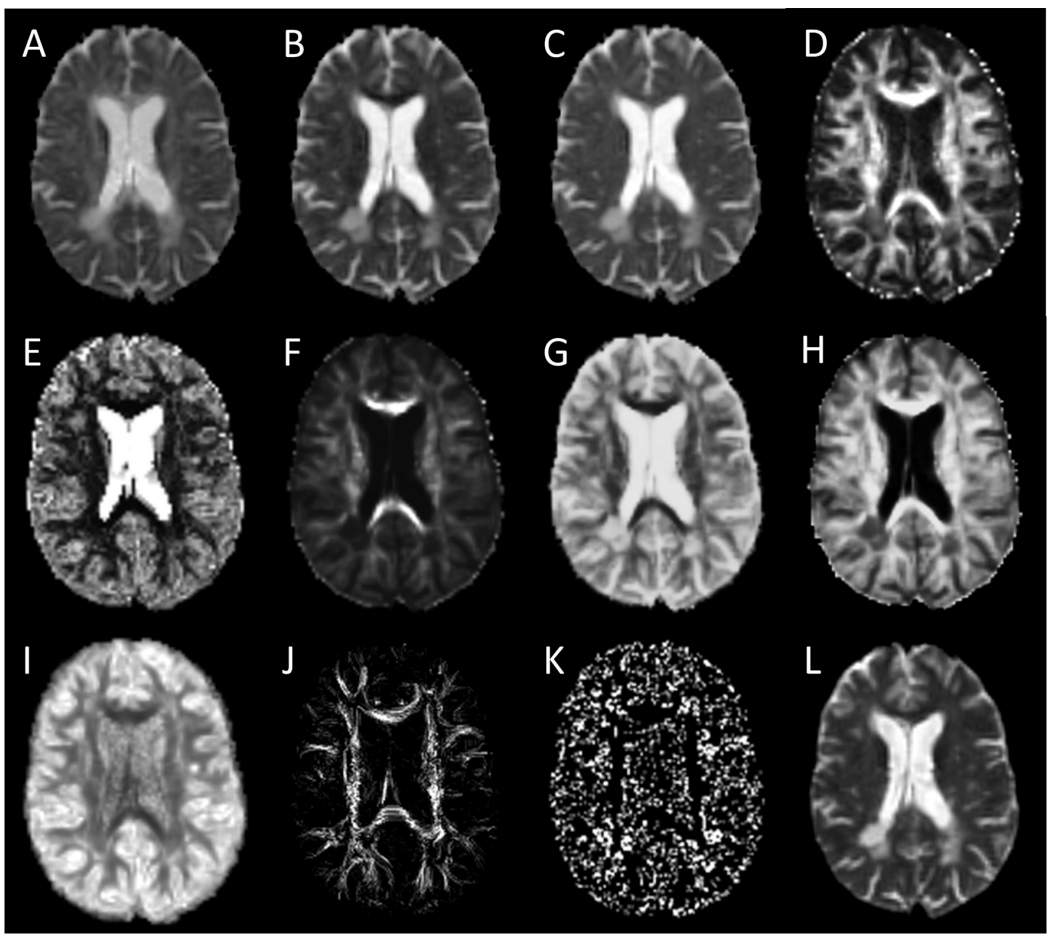


FIGURE 2 | Sample diffusion metrics based on DTI, ssHARDI, and DTI tractography models. Shown are measures from DTI: **(A)** axial diffusivity, **(B)** radial diffusivity, **(C)** mean diffusivity, and **(D)** fractional anisotropy; ssHARDI: **(E)** orientation dispersion index, **(F)** density, **(G)** diameter, **(H)** intracellular volume fraction, and **(I)** orientation distribution function (ODF) energy; and DTI tractography: **(J)** fiber density index and **(K)** fiber termination index. Shown in **(L)** is an example b0 image from DTI for reference.

compare outcomes from different datasets. This included: (1) single-shell, 2-shell, and 3-shell HARDI data freely available online from the Human Connectome Project (HCP) (Van Essen et al., 2012); and (2) single-shell data from our own study. The comparisons were done both visually and quantitatively with a concentration on 3 ssHARDI measures out of AMICO ActiveAX: axonal diameter, axonal density, and ICVF. NODDI ODI had been shown to be similar between calculations of

single- and multi-shell data (Zhang et al., 2012). In quantitative assessment, we computed the variance of the aforementioned measures at a regional level in 10 example white matter structures bihemispherically: forceps minor, forceps major, genu and splenium of the corpus callosum, and posterior limb of the internal capsule. Each ROI was sized 4 × 4 pixels, which was the maximal uniform dimension that could be fitted within these structures. Statistical analyses were conducted using one-way ANOVA followed by correction for multiple comparisons where applicable.

DTI Tractography Analysis

We performed tractography based on DTI using the Diffusion Toolkit (MGH GCRC, United States) (Wang and Wedeen, 2007), which applied a deterministic algorithm, fiber assignment by continuous tracking (FACT) (Wang and Wedeen, 2007). Streamline propagation followed a 35° angular threshold as recommended.

Using the rigid-body transformation matrix derived above from FSL, we also aligned the tractography to the T1 structural space. Based on the TrackVis method (v0.6.1) (Wang and Wedeen, 2007), we evaluated two main tractography measures (see **Figure 2**): streamline density index (FDi) and streamline endpoint index (FTi), indicating the counts of streamlines passing through or terminating in each voxel, respectively. Our in-house experiment showed that tractography features based on DTI were similar to ssHARDI, so we focused on DTI only here.

SVM Analysis

In feature ranking, we split the whole dataset randomly into 10 folds (portions), using 9 folds at a time, and repeated 10 times. In classifications, 10-fold cross-validation was performed, with 9 folds for training, and the 10th held-out fold for testing, in each iteration. As a standard practice, we also normalized all feature outcomes to the range 0–1.

We applied a linear SVM to rank the diffusion features through recursive feature elimination (SVM -RFE) (Guyon et al., 2002), with the regularization term fixed to one for optimal model generalization. The squared weights of each attribute of the SVM served as the ranking criteria such that a feature with the smallest weight ranked the lowest. The average ranking of a feature from all model constructions (100 in total) represented the final ranking of the feature.

To assess the contribution of each feature to model classifications, we developed another linear SVM with iterative construction (SVM-IC), achieved by adding features to the model one-by-one from the highest to the lowest rankings. With each feature added, the model is reconstructed through 10-fold cross validation. This provided data to build the rank aggregated receiver operating characteristic (ROC) curves for each combination of successively ranked features, and compute evaluation metrics including the area under the ROC (AUROC) and accuracy, averaged from all tests.

Patient Stratification

To explore how the imaging features could differentiate patient groups, we stratified the participants as having high and low number of MS lesions as a surrogate of disease activity. The lesion numbers from individual patients were ranked by percentiles. To maximize the likelihood of detecting potential differences, our analysis focused on two patient groups that had highly different lesion numbers: one with the most lesions ($\geq$35) that ranked above the 75%^{ile}, and the other with the least lesions ($\leq$15) that ranked below the 25%^{ile}. Group comparisons were done for both lesion tissue (whole lesion ROIs) and NAWM regions, using similar SVM strategies as described above for feature ranking and tissue classification.

Statistical Analysis

Assessment of all outcomes used the Scikit-Learn package in Python (v0.22.1) and R (v3.6.3). Comparing model performance used the McNemar's test (Agresti, 1990) for accuracy, and DeLong's test (DeLong et al., 1988) for AUROC, including model against chance, and between models successively generated.

RESULTS

Sample Characteristics

Of the 52 participants, the age range was 18–60 years, Expanded Disability Status Scale was 0–5.5; 36 were women. In total, we identified 2139 lesion ROIs, 4 to 169 per subject; 1560/2139 lesions had matching contralateral NAWM ROIs, 1 to 119 per subject. Among the 1560 lesions, 243 had core-shell analysis (**Table 1**). In addition, 13 participants had $\geq$ 35 lesions, totaling 818 lesions, and 12 participants had $\leq$ 15 lesions, totaling 119 lesions. Similarly, there were 818 and 119 matching contralateral NAWM ROIs in each patient group respectively. Further, example outcomes from ssHARDI modeling appeared similar to that from multi-shell data, including the measures from our own DTI scans (**Figure 3**). Quantitatively, there were no significant differences in variance for diameter, density, or ICVF ($p = 0.75$, 0.18, and 0.11, respectively) between the different shell calculations using HCP data or data from our own study (**Table 2**). Further exploration using paired Student's t-tests showed that there were also no significant differences ($p > 0.05$) between the 1-shell measures from HCP and our DTI data in variance of any of the assessed variables following ssHARDI modeling. In total, there were 11 features calculated from all diffusion models.

TABLE 1 | Lesion counts and volumes by group.

Lesion ROIs	Number	Lesion volume (mm^3)			
		Mean	Standard error	Min	Max
All	2139	256.21	25.65	3.81	16443.64
NAWM-paired	1560	45.41	1.86	3.81	929.88
Core-shell	243	158.14	8.34	35.29	929.88
Shell	–	144.89	6.93	34.33	735.35
Core	–	13.25	1.79	0.95	243.20

ROI, regions of interest; NAWM, normal appearing white matter.

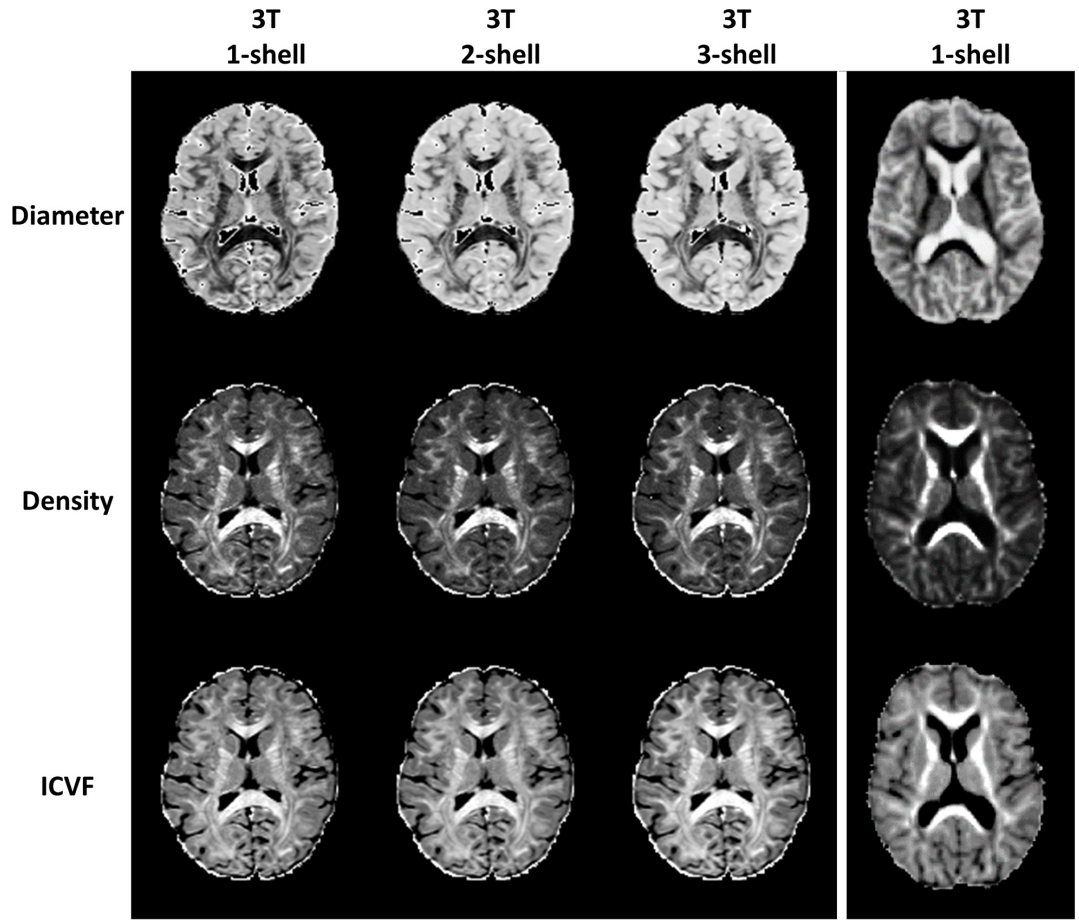


FIGURE 3 | Outcome comparison between single-shell and multi-shell acquisitions. Shown are sample HARDI outcomes based on the ActiveAx model implemented in AMICO using the online Human Connectome Project data (left columns): $b = 1000$, $b = 1000$ and 2000, and $b = 1000$, 2000, and 3000 for the 1-, 2-, and 3-shell, respectively; and using our own diffusion data in this study based on ssHARDI (right column), with $b = 1000$. All represent *in vivo* datasets.

TABLE 2 | Comparison of variance between measures from different shells and datasets computations.

	HCP1	HCP2	HCP3	In-house	ANOVA *p*-value	ssHARDI *p*-value
Diameter	1.416713	1.313007	1.449823	0.939283	0.749132	0.131976
Density	3.62E-06	2.94E-06	3.69E-06	8.1E-06	0.175857	0.192801
ICVF	0.006766	0.006788	0.006341	0.001743	0.114337	0.050831

HCP1-3, Human connectome project data from 1 to 3 shell acquisitions; In-house, our own "1-shell" DTI data; ssHARDI p-value, Student's t-tests for variances between HCP1 and In-house data.

Lesion-NAWM Analysis

Tissue alignment metrics ranked higher than magnitude metrics. Specifically, the top three rankings were: HARDI ODF energy, DTI FA, and HARDI ODI (**Figure 4**). Tractography FDi (fourth), DTI AD (fifth), and tractography FTi (sixth) all ranked within the top half of the 11 features, relatively higher than the other DTI and ssHARDI features.

Further assessments using the classification model revealed similar trends. Essentially, combining all top three features (ODF energy, FA, and ODI) in the SVM-IC model achieved a 0.65 accuracy and 0.71 AUROC (**Figures 5** and **6**). McNemar tests showed that model accuracy with ODF energy alone significantly outperformed chance ($p < 2.2e\text{-}16$) and improved further with addition of FA ($p < 2.2e\text{-}16$). ROC analysis mirrored these results. The AUROC was significantly better using ODF energy alone ($p = 2.95e\text{-}6$) than chance and improved with FA ($p < 2.2e\text{-}16$). Further, the AUROC peaked at 0.71 with both ODI and FDi added, but did not change significantly with further inclusion of the remaining features (**Table 3**). Model accuracy peaked at 0.66 with all but the lowest-ranked feature (DTI RD) included.

Core-Shell Analysis

Feature ranking in this assessment used 9 of the 11 diffusion features, with tractography FDi and FTi excluded due to their

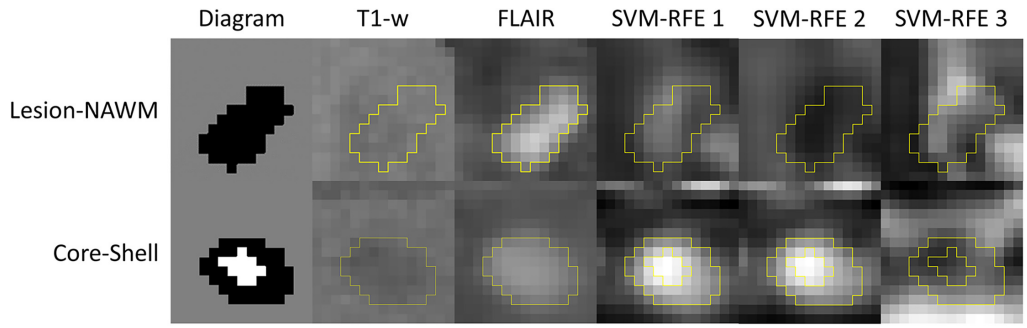


FIGURE 4 | Example lesion regions and top diffusion metrics identified in the tissue separation processes. Shown are diagrams for a whole lesion (black, top) and a core-shell lesion (white and black, bottom) used for the lesion versus NAWM and core versus shell classifications, as well as corresponding appearances in the T1-weighted and FLAIR MRI (columns 1–3). The other columns (4–6) demonstrate the top three diffusion features selected by the recursive feature elimination algorithm (SVM-RFE) in respective classification tasks, which are orientation distribution function energy, fractional anisotropy, and orientation dispersion index (top); and mean diffusivity, radial diffusivity, and fractional anisotropy (bottom).

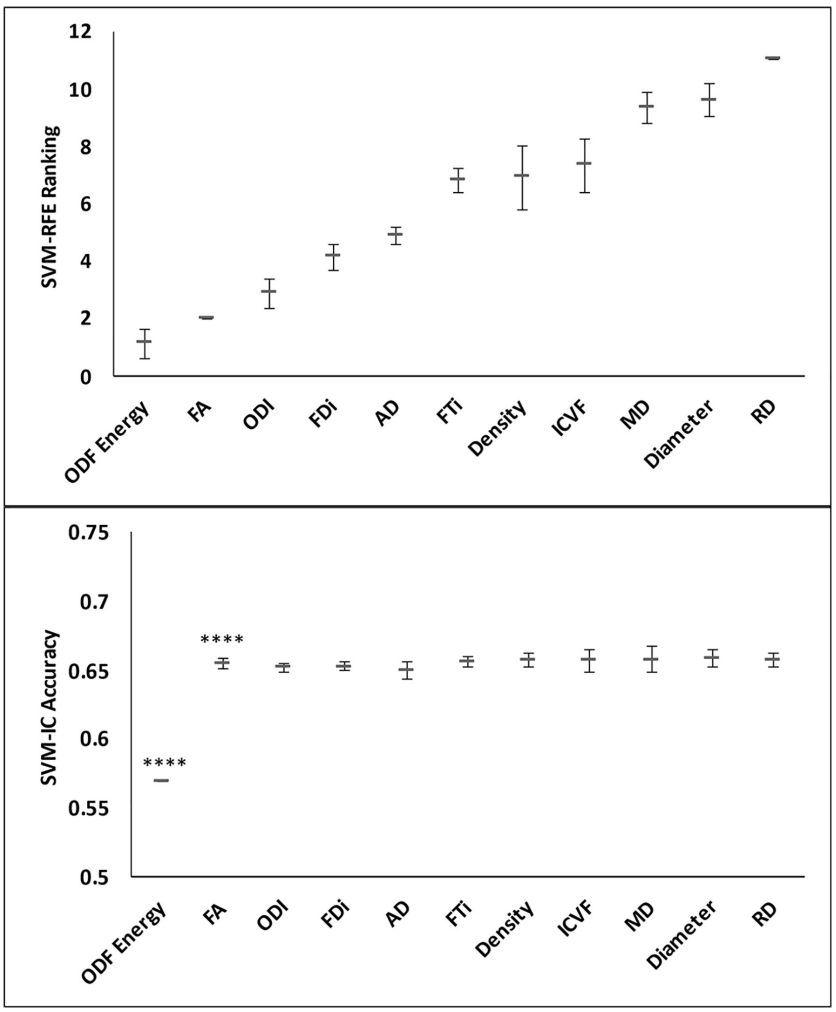


FIGURE 5 | Feature ranks and classification accuracy in lesion-NAWM analysis. **Top panel:** The mean (standard deviation) rankings of the 11 diffusion features based on the linear recursive feature elimination (RFE) algorithm in SVM. **Bottom panel:** Accuracy of the classification models (SVM-IC) obtained by adding features one at a time, starting from the highest to the lowest rankings. The stars indicate features that contribute to significant improvement in classification accuracy (****$p < 0.0001$).

TABLE 3 | The AUROC outcome of the linear SVM classification models.

Features	Lesion vs. NAWM		Features	Core vs. Shell	
	AUROC	stdev		AUROC	stdev
ODF energy	0.568	0.030	MD	0.594	0.080
FA	0.707	0.023	RD	0.597	0.088
ODI	0.710	0.026	FA	0.598	0.092
FDi	0.713	0.026	ODF energy	0.600	0.083
AD	0.710	0.029	AD	0.594	0.081
FTi	0.711	0.028	ICVF	0.595	0.081
Density	0.710	0.028	Density	0.598	0.082
ICVF	0.711	0.028	ODI	0.595	0.086
MD	0.710	0.028	Diameter	0.593	0.086
Diameter	0.710	0.028			
RD	0.710	0.028			

AUROC, area under the receiver operating characteristic curve; stdev, standard deviation.

sparse representation in relatively small ROIs. The top three ranked metrics were: DTI MD, RD, and FA. HARDI ODF energy and DTI AD followed in ranking, slightly better than the other HARDI measures, which ranked at 6th–9th (ICVF, Density, ODI, and Diameter).

Classification analysis further revealed the importance of top-ranked metrics in core-shell analysis. The highest-ranked feature, MD, alone accounted for a classification accuracy of 0.59, significantly greater than chance (p = 4.98e-5; **Figures 7** and **8**). Adding features up to the fifth one (AD), the classification accuracy peaked, at 0.60 (p = 3.86e-6). In addition, the model with MD alone achieved an AUROC of 0.59, which improved to the peak at 0.60 when combining with the top 2nd to 4th features (RD, FA, ODF energy), but the AUROC values were not significantly different from chance (p = 0.088; **Table 3**).

Lesions and NAWM Analyses Between Subjects

These analyses used all of the 11 diffusion features as applied in assessing whole lesion pathology. In comparing the lesion tissue between subjects who had 75%^{ile} high versus 25%^{ile} low lesion load, FA ranked the highest, which alone had an accuracy of 0.589 and AUROC of 0.620. Combining FA with the 2nd–4th features, AD, ICVF, and Diameter, slightly increased the accuracy, which reached the peak at 0.605; further addition of FTi as the fifth feature led to the peak AUROC of 0.638, but no models performed significantly better than chance (p = 0.200) in this analysis. In contrast, ODF energy ranked the best in classifying the matching NAWM regions between the same patient groups. Specifically, ODF energy alone achieved a near-peak accuracy of 0.616 and AUROC of 0.662, considerably better than chance (p = 0.020). Adding the second best parameter, neurite density, achieved the peak values for accuracy at 0.617 and for AUROC at 0.662, almost identical to the ODF energy alone results. There was no further improvement with addition of any other parameters (**Figures 9** and **10**).

DISCUSSION

Using commonly available dMRI data, we showed the feasibility of conducting ssHARDI analysis and the complementary value of parameters from relatively simple and complex diffusion models for assessing MS pathology. It appears that tissue alignment and orientation measures from ssHARDI and DTI are particularly sensitive to the existence of lesions in a patient, and to subtle structural differences between NAWM areas (specifically the ODF energy metric) between patients with high versus low disease activity. In contrast, tissue diffusivity and alignment metrics from DTI are the best in distinguishing intra-lesion pathology, and in separating lesion activity between high and low disease patients, especially MD and FA.

The ability to evaluate sub-voxel-based tissue properties through ssHARDI modeling would be important in a clinical setting. However, ssHARDI studies often face the challenge of not obtaining enough diversity of measures for evaluation, unlike multi-shell models. To compensate, a prior study applied a partial, 2-compartment model and showed that both neurite density and dispersion indices from ssHARDI decreased in the NAWM of MS patients, and the reduction in the right internal capsule correlated with MS duration (Magnollay et al., 2014). In the present study, we undertook an alternative approach with the assistance of an efficient modeling method, AMICO. This provided several sub-voxel measures of neurite diameter, density, and dispersion. Both visual and quantitative analyses show that the variances at a ROI level are equivalent between

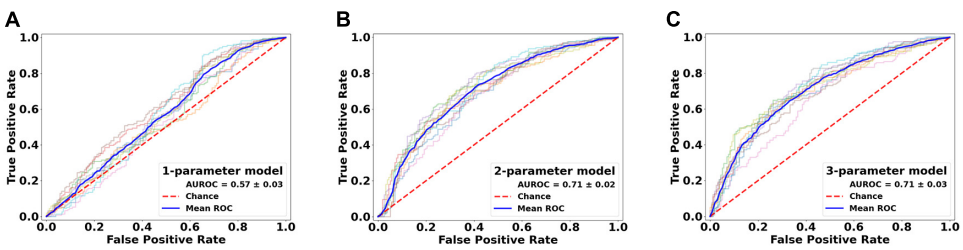


FIGURE 6 | Performance comparison of the classification models in lesion versus NAWM analysis. Shown are the ROC curves for models constructed from the top three ranked diffusion features: **(A)** orientation distribution function (ODF) energy alone, **(B)** ODF energy + fractional anisotropy (FA), and **(C)** ODF energy + FA + orientation dispersion index. ROC, receiver operating characteristics; AUROC, area under the ROC curve.

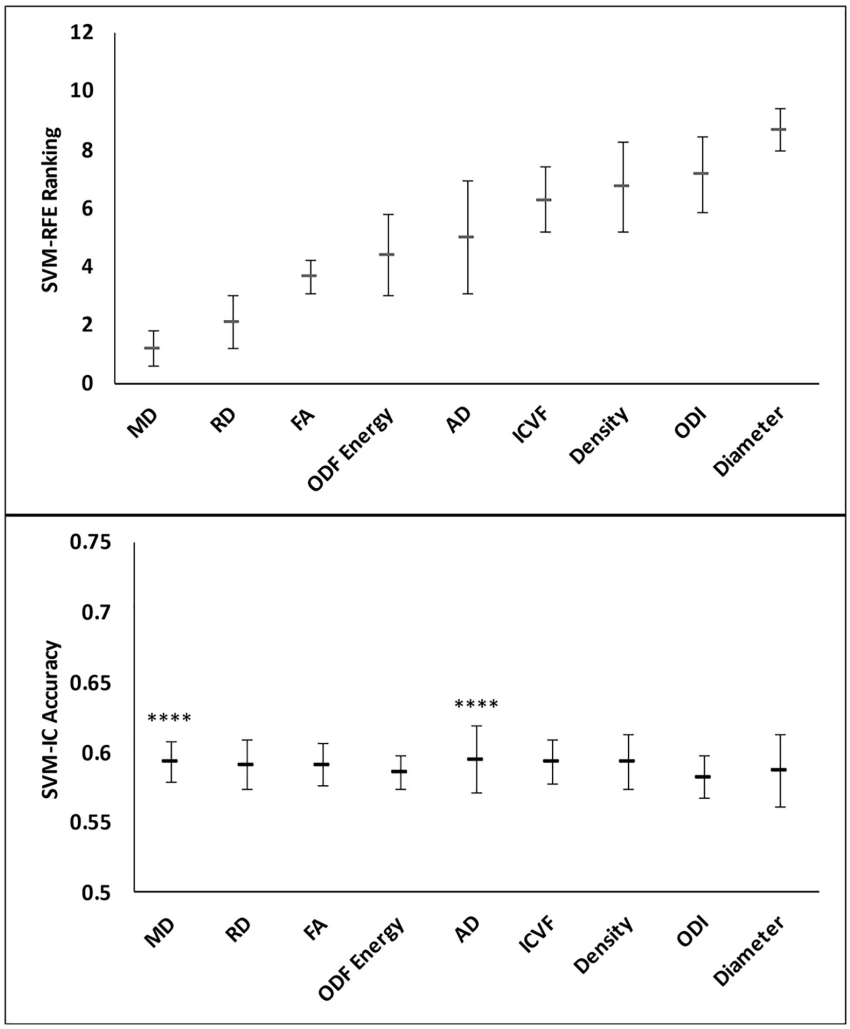


FIGURE 7 | Feature ranks and classification accuracies in lesion core-shell analysis. **Top Panel:** The mean (standard deviation) rankings of the nine diffusion features used in this analysis based on the linear recursive feature elimination (RFE) algorithm in SVM. **Bottom Panel:** Accuracy of the classification models (SVM-IC) achieved by adding features one at a time, starting from the highest to the lowest rankings. The stars indicate features that contribute to significant improvement in classification accuracy (****$p < 0.0001$).

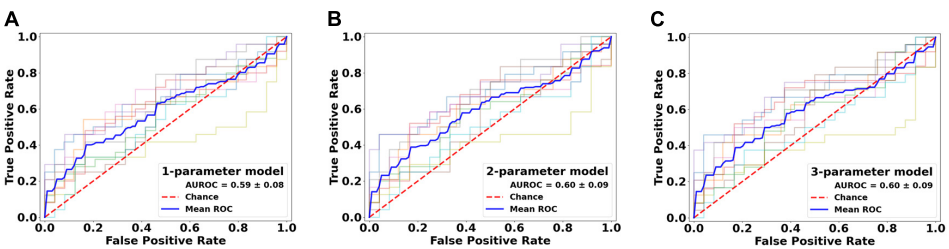


FIGURE 8 | Performance comparison of the classification models in the lesion core versus shell analysis. Shown are the ROC curves for models constructed from the top three ranked diffusion features: **(A)** mean diffusivity (MD) alone, **(B)** MD + radial diffusivity (RD), and **(C)** MD + RD + fractional anisotropy. ROC, receiver operating characteristics; AUROC, area under the ROC curve.

these measures derived using either 1-shell or multi-shell data, and using either online or own DTI data in ssHARDI modeling. In addition, based on the distribution of diffusion at all sampling directions in a voxel, we also created a new measure of tissue organization, ODF energy, to increase the capacity of ssHARDI.

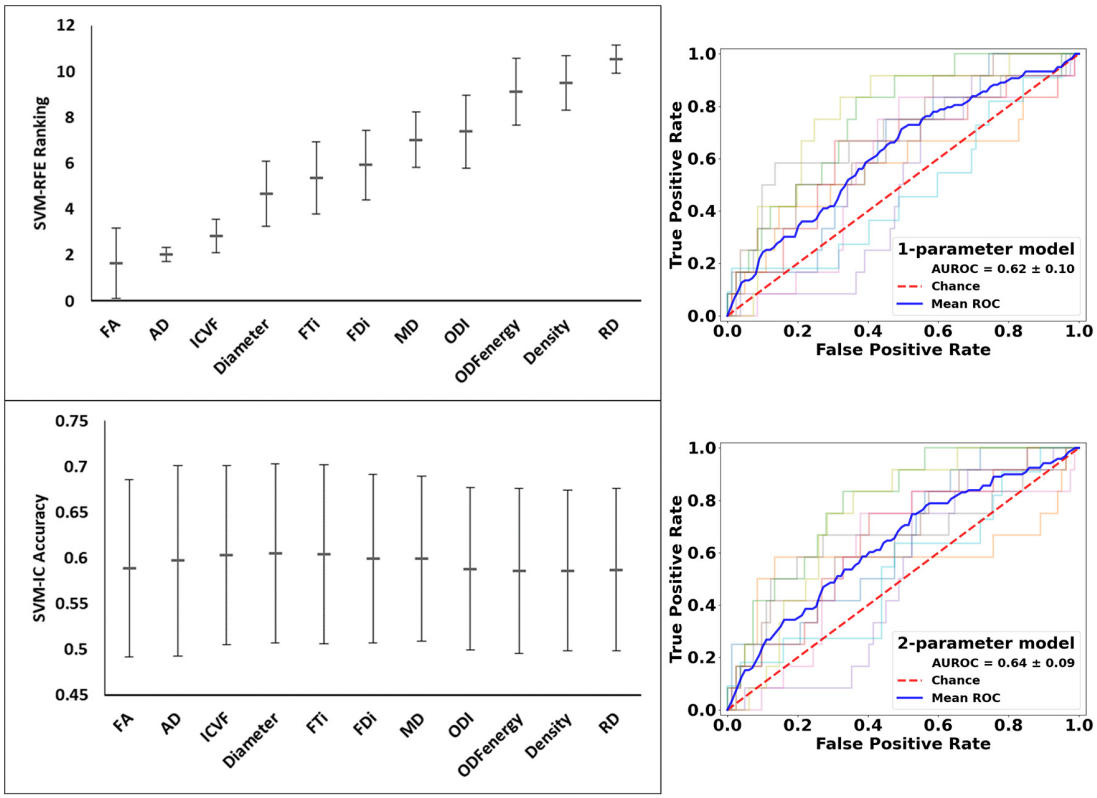


FIGURE 9 | Feature ranks and performance comparison in the classification of lesion tissue between high and low disease patients. **Top Left Panel:** The mean (standard deviation) rankings of the 11 diffusion features used in this analysis based on the linear recursive feature elimination (RFE) algorithm in SVM. **Bottom Left Panel:** Accuracy of the classification models (SVM-IC) achieved by adding features one at a time, starting from the highest to the lowest rankings. **Right Panel:** ROC curves for models constructed using the top two ranked diffusion features: (top fractional anisotropy (FA) alone, and bottom) FA + axial diffusivity (AD). ROC, receiver operating characteristics; AUROC, area under the ROC curve.

To maximize the understanding of diffusion metrics in tissue assessment, our study considered several modeling approaches, 3 regarding ssHARDI alone: ActiveAx, NODDI, and diffusion ODF. Multi-model analysis is critical for assessing complex pathologies as observed in MS (Reich et al., 2018). However, this strategy also led to a large volume of parameters. A caveat of multiparametric approaches (Cercignani and Bouyagoub, 2018) is the difficulty of assigning importance to parameters that have overlapping sensitivities. Here we took advantage of SVM, a well-recognized machine learning method. The SVM-RFE has shown to be robust to feature redundancy and model overfitting, and in a linear form, the SVM can assign unique coefficients to each parameter, thereby identifying the specific contribution of a parameter to individual classification tasks (Guyon et al., 2002).

In lesion–NAWM analysis, nearly all top rankings were tissue orientation and alignment metrics, particularly the top 3: ODF energy, FA, and ODI. Based on definition, ODF energy measures the orientational complexity of a structure. The higher the value, the more misalignment in the structure. FA detects changes in both axonal density and alignment anisotropy (Zhang et al., 2012), and therefore is partially explainable by ODI. The leading performance of these orientation features in this analysis may indicate that the most critical changes following lesion formation in MS are tissue damage, such as inflammatory demyelination and axonal injury. A direct consequence of this pathology is increased structural heterogeneity, rather than simple alterations in neurite density. This is consistent with prior findings showing decreased FA in nearly all lesion studies in MS compared to the NAWM (Schmierer et al., 2007; Inglese and Bester, 2010), and ODI changes with lesion formation (Grussu et al., 2017; Schneider et al., 2017) and repair (Luo et al., 2019), although the consistency of ODI changes deserves further validation (Spano et al., 2018). Notably, optimal classification between lesion and NAWM ROIs in the current study required the addition of FA. Previously, tractography FDi also showed strong correlations with FA (Roberts et al., 2005; Ukmar et al., 2012), supporting the relationship between these highly ranked features. The unique role of tissue alignment and orientation measures may facilitate early detection and even prediction of lesion pathology in future studies, promoting early management.

The core-shell analysis was designed to probe the intra-lesion patterns of pathology as seen in chronic MS lesions. When active, these lesions show demyelinated, hypocellular cores and inflammatory demyelinating shells; while inactive, present with hypocellularity with no active demyelination in lesion territories

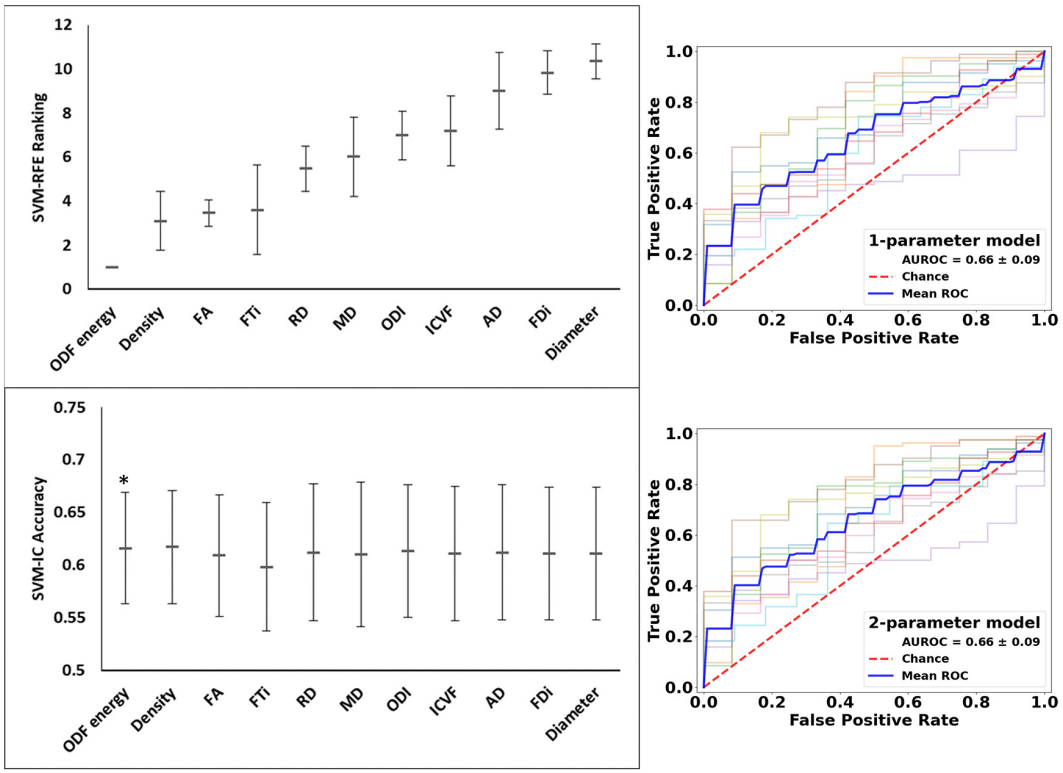


FIGURE 10 | Feature ranks and performance comparison in the classification of NAWM tissue between high and low disease patients. **Top Left Panel:** The mean (standard deviation) rankings of the 11 diffusion features used in this analysis based on the linear recursive feature elimination (RFE) algorithm in SVM. **Bottom Left Panel:** Accuracy of the classification models (SVM-IC) achieved by adding features one at a time, starting from the highest to the lowest rankings. **Right Panel:** ROC curves for models constructed using the top two ranked diffusion features: [top orientation distribution function (ODF) energy alone, and bottom] ODF energy + Density (AD). ROC, receiver operating characteristics; AUROC, area under the ROC curve (*$p < 0.05$).

(Pittock and Lucchinetti, 2007). As such, the expected differences between core and shell are the degree of cellularity, where the hypocellular core should have higher diffusivity than the cell-rich, inflammatory shell. Indeed, the 3 DTI measures (MD, RD, and FA) ranked the highest in our core-shell analysis, and MD played a dominant role. While there is lack of core-shell studies using dMRI in the literature, prior evidence attests the sensitivity of MD to subtle changes in MS pathology. One report showed that MD increased 5 months before the occurrence of active MS lesions, and pre-lesion MD correlated significantly with the MD of the lesions 2 months after active inflammation (Liu et al., 2012). In the present study, the lack of significance of the MD model in AUROC compared to chance may be due to several reasons, including the small number of such lesions and their heterogeneity in pathology. For instance, chronic inactive lesions may not show significant core-shell differences, and their inclusion may have caused artifacts. Alternatively, robust analysis of core-shell activity in chronic active lesions can improve our understanding of the ongoing pathology and so disease progression in MS patients (Giorgio et al., 2014).

Our inter-patient analysis results appear to agree with the findings described above. Between patients who have high versus low disease activity, FA ranked the highest in lesion tissue classifications; while in NAWM classifications, ODF energy was the best. In general, T2 MRI lesion load reflects disease activity, and in a long-term, it correlates with patient disability in MS as seen after 20 years of disease onset (Fisniku et al., 2008). Therefore, patients with higher lesion load are expected to have greater tissue damage, likely affecting both neurite density and orientation, and so greater changes in lesion FA between the patient groups. On the other hand, increased disease activity would also implicate increased pathology in the lesion-free areas, such as the NAWM, yet the changes wherein should be much less than in the visible plaques. Indeed, our NAWM analysis suggests that the pattern of differences between high and low disease patients is mainly orientation based, as reflected by ODF energy, deserving further confirmation.

Collective observations of this study may suggest that MD and FA are important diffusion metrics for assessing MS pathology. In particular, they may serve as top options in core versus shell analysis, disease severity comparisons between patients (e.g., lesion FA), and at a lesser extent, lesion versus non-lesion analysis. However, parameters from the relatively complex models appear to be more competitive in NAWM-associated evaluations than the DTI indices, particularly the orientation-driven metrics (ODF energy and ODI), suggesting the value of multi-model analysis. Currently, dMRI is one of the unique imaging approaches that can potentially evaluate both myelin and

axonal properties *in vivo*. However, a variety of other techniques are underway in this regard, including those for myelin mapping using either advanced or conventional MRI. For example, myelin estimation based on simultaneous T1/T2 relaxometry and proton density mapping correlated strongly with histological myelin as seen in post-mortem MS brain samples (Ouellette et al., 2020). Another study demonstrated that quantitative susceptibility provided additional myelin information in brain NAWM of MS patients, independent of FA and RD (Yu et al., 2019). Further, pre-operative myelin mapping using T1/T2 ratio showed the potential to predict outcomes of trigeminal neuralgia following Gamma knife radiosurgery, while FA and RD demonstrated similar values in this prediction as estimators of pre-operative axons (Li et al., 2020). These findings further support the importance of multi-dimensional analysis in MS and similar diseases, including combining multi-model dMRI, and various other candidate metrics of myelin and axons.

We note a few limitations in this study. The spatial resolution of our dMRI was moderate. However, we registered all diffusion outcomes to the high-resolution T1 MRI to mitigate the impact, and our ssHARDI maps appeared similar to those obtained using high resolution dMRI from the HCP both visually and quantitatively. Next, to optimize data quality, we excluded lesions that did not have a "clean" match of contralateral NAWM regions, and lesions not large enough for core-shell analysis, limiting the sample size. Nonetheless, SVM analyses enabled determination of critical parameters, so the risk of model overfitting is minimal. Finally, there were lesions that appeared overlapping and they were segmented as single confluent regions. Given the inhomogeneity of signal intensity in such lesions, it may have somewhat affected the lesion-NAWM results (e.g., underestimation of tissue differences). However, we applied ROI means in all quantitative analyses, which should have minimized the effect, if any. The impact on core-shell analysis is expected to be less than the above, as regions of inhomogeneity may be included in both lesion portions. In the future, we seek to use higher resolution DTI and evaluate patients with different types of diseases and lesion pathologies to confirm our findings, including datasets with histology-verified lesion activity. Additionally, we also plan to investigate the utility of the top-performing diffusion features in assessing pre-lesion NAWM pathologies, and in predicting the occurrence of MS lesions, to promote clinical applications.

Collectively, using commonly available dMRI data, it is possible to perform competitive ssHARDI modeling. Combining machine learning with robust ssHARDI and DTI metrics may provide advanced assessment of lesion and NAWM pathology, including mean diffusivity for intra-lesion pathology in MS and similar diseases.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Conjoint Health Research Ethics Board, the University of Calgary. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

OO participated in study design, data analysis and interpretation, and manuscript drafting and editing. W-QL participated in study design, data acquisition and analysis, and manuscript editing. BP, MK, and LM participated in study design, data acquisition, and manuscript editing. YZ participated in study design, data acquisition and interpretation, and manuscript drafting and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Alberta Innovates and Multiple Sclerosis Society of Canada for acquiring study resources, operating the study, and disseminating results. This research was also benefited from personnel awards from the Alberta Innovates and Canadian Network for MS Clinics (W-QL); and scholarship support from the NSERC CREATE I3T program, Bonvicini Neuroscience Scholarship, the Government of Alberta Queen Elizabeth II Graduate Scholarship, and Alberta Graduate Education Scholarship (OO).

ACKNOWLEDGMENTS

We thank the patient volunteers for participating in the clinical study and the MS clinical research team for supporting the study. Data were obtained in part from the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis. Alan Agresti*. New York, NY: Wiley.

Alexander, D. C., Hubbard, P. L., Hall, M. G., Moore, E. A., Ptito, M., Parker, G. J., et al. (2010). Orientationally invariant indices of axon diameter and density from diffusion MRI. *Neuroimage* 52, 1374–1389. doi: 10.1016/j.neuroimage.2010.05.043

Andersson, J. L. R., and Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* 125, 1063–1078. doi: 10.1016/j.neuroimage.2015.10.019

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

Battaglini, M., Jenkinson, M., and De Stefano, N. (2012). Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum. Brain Mapp.* 33, 2062–2071. doi: 10.1002/hbm.21344

Cercignani, M., and Bouyagoub, S. (2018). Brain microstructure by multi-modal MRI: Is the whole greater than the sum of its parts? *Neuroimage* 182, 117–127. doi: 10.1016/j.neuroimage.2017.10.052

Daducci, A., Canales-Rodríguez, E. J., Zhang, H., Dyrby, T. B., Alexander, D. C., and Thiran, J.-P. (2015). Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data. *Neuroimage* 105, 32–44. doi: 10.1016/j.neuroimage.2014.10.026

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837. doi: 10.2307/2531595

Descoteaux, M. (2015). "High angular resolution diffusion imaging (HARDI)," in *Wiley Encyclopedia of Electrical and Electronics Engineering*, (Hoboken, NJ: John Wiley & Sons, Inc), 1–25. Available online at: http://doi.wiley.com/10.1002/047134608X.W8258 (accessed November 05, 2020).

Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R. (2007). Regularized, fast, and robust analytical Q-ball imaging. *Magn. Reson. Med.* 58, 497–510. doi: 10.1002/mrm.21277

Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., et al. (2008). Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain* 131, 808–817. doi: 10.1093/brain/awm329

Giorgio, A., Stromillo, M. L., Bartolozzi, M. L., Rossi, F., Battaglini, M., De Leucio, A., et al. (2014). Relevance of hypointense brain MRI lesions for long-term worsening of clinical disability in relapsing multiple sclerosis. *Mult. Scler* 20, 214–219. doi: 10.1177/1352458513494490

Grussu, F., Schneider, T., Tur, C., Yates, R. L., Tachrount, M., Ianuș, A., et al. (2017). Neurite dispersion: a new marker of multiple sclerosis spinal cord pathology? *Ann. Clin. Transl. Neurol.* 4, 663–679. doi: 10.1002/acn3.445

Grussu, F., Schneider, T., Zhang, H., Alexander, D. C., and Wheeler-Kingshott, C. A. M. (2014). Single-shell diffusion MRI NODDI with in vivo cervical cord data. *Proc. Intl. Soc. Mag. Reson. Med.* 22:1716.

Guyon, I., Weston, J., and Barnhill, S. (2002). Gene selection for cancer classification using support vector machines. *Mach Learn.* 46, 389–422. doi: 10.1023/A:1012487302797

Inglese, M., and Bester, M. (2010). Diffusion imaging in multiple sclerosis: research and clinical implications. *NMR Biomed* 23, 865–872. doi: 10.1002/nbm.1515

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med Image Anal.* 5, 143-156. doi: 10.1016/S1361-8415(01)00036-6

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Klistorner, A., Wang, C., Yiannikas, C., Parratt, J., Dwyer, M., Barton, J., et al. (2018). Evidence of progressive tissue loss in the core of chronic MS lesions: a longitudinal DTI study. *NeuroImage Clin.* 17, 1028–1035. doi: 10.1016/j.nicl.2017.12.010

Li, C. M. F., Hung, P. S. P., Chu, P. P. W., Tohyama, S., and Hodaie, M. (2020). Trigeminal neuralgia associated with multiple sclerosis: a multimodal assessment of brainstem plaques and response to gamma Knife radiosurgery. *Mult. Scler J.* 26, 1877–1888. doi: 10.1177/1352458519886070

Liu, Y., Mitchell, P. J., Kilpatrick, T. J., Stein, M. S., Harrison, L. C., Baker, J., et al. (2012). Diffusion tensor imaging of acute inflammatory lesion evolution in multiple sclerosis. *J. Clin. Neurosci.* 19, 1689–1694. doi: 10.1016/j.jocn.2012.03.022

Luo, T., Oladosu, O., Rawji, K. S., Zhai, P., Pridham, G., Hossain, S., et al. (2019). Characterizing structural changes with devolving remyelination following experimental demyelination using high angular resolution diffusion MRI and texture analysis. *J. Magn. Reson. Imaging* 49, 1750–1759. doi: 10.1002/jmri.26328

Magnollay, L., Grussu, F., Wheeler-Kingshott, C. A. M. M. M., et al. (2014). "An investigation of brain neurite density and dispersion in multiple sclerosis using single shell diffusion imaging," in *Proceedings of the International Society for Magnetic Resonance in Medicine (Milan)*, 2048.

Msif Tmsif. (2020). *Atlas of MS*. Available online at: www.atlasofms.org (accessed November 05, 2020).

Ouellette, R., Mangeat, G., Polyak, I., Warntjes, M., Forslin, Y., Bergendal, Å, et al. (2020). Validation of rapid magnetic resonance myelin imaging in multiple sclerosis. *Ann. Neurol.* 87, 710–724.

Pittock, S. J., and Lucchinetti, C. F. (2007). The pathology of MS. *Neurologist* 13, 45–56. doi: 10.1097/01.nrl.0000253065.31662.37

Reich, D. S., Lucchinetti, C. F., and Calabresi, P. A. (2018). Multiple sclerosis. *N. Engl. J. Med.* 378, 169–180.

Roberts, T. P. L., Liu, F., Kassner, A., Mori, S., and Guha, A. (2005). Fiber density index correlates with reduced fractional anisotropy in white matter of patients with glioblastoma. *AJNR Am. J. Neuroradiol.* 26, 2183–2186.

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., et al. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *Neuroimage* 59, 3774–3783. doi: 10.1016/j.neuroimage.2011.11.032

Schmierer, K., Wheeler-Kingshott, C. A. M. M., Boulby, P. A., Scaravilli, F., Altmann, D. R., Barker, G. J., et al. (2007). Diffusion tensor imaging of post mortem multiple sclerosis brain. *Neuroimage* 35, 467–477. doi: 10.1016/j.neuroimage.2006.12.010

Schneider, T., Brownlee, W., Zhang, H., Ciccarelli, O., Miller, D. H., and Wheeler-Kingshott, C. G. (2017). Sensitivity of multi-shell NODDI to multiple sclerosis white matter changes: a pilot study. *Funct. Neurol.* 32, 97–101. doi: 10.11138/fneur/2017.32.2.097

Spano, B., Giulietti, G., Pisani, V., Morreale, M., Tuzzi, E., Nocentini, U., et al. (2018). Disruption of neurite morphology parallels MS progression. *Neurol. Neuroimmunol. NeuroInf.* 5:e502. doi: 10.1212/nxi.0000000000000502

Taquet, M., Scherrer, B., Boumal, N., Macq, B., and Warfield, S. K. (2013). Estimation of a multi-fascicle model from single B-value data with a population-informed prior. *Med. Image Comput. Comput. Assist. Interv.* 16(Pt. 1), 695–702. doi: 10.1007/978-3-642-40811-3_87

Tournier, J. D. (2019). Diffusion MRI in the brainTheory and concepts. *Prog. Nucl. Magn. Reson. Spectrosc.* 11, 1–16.

Ukmar, M., Montalbano, A., Makuc, E., Specogna, I., Bratina, A., Longo, R., et al. (2012). Fiber density index in the evaluation of the spinal cord in patients with multiple sclerosis. *Radiol. Med.* 117, 1215–1224. doi: 10.1007/s11547-012-0848-0

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E., Bucholz, R., et al. (2012). The human connectome project: a data acquisition perspective. *Neuroimage* 62, 2222–2231.

Wang, R., and Wedeen, V. J. (2007). Diffusion toolkit: a software package for diffusion imaging data processing and tractography. *Proc. Intl. Soc. Mag. Reson. Med.* 15:3720.

Yu, F. F., Chiang, F. L., Stephens, N., Huang, S. Y., Bilgic, B., Tantiwongkosi, B., et al. (2019). Characterization of normal-appearing white matter in multiple sclerosis using quantitative susceptibility mapping in conjunction with diffusion tensor imaging. *Neuroradiology* 61, 71–79. doi: 10.1007/s00234-018-2137-7

Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., and Alexander, D. C. N. O. D. D. I. (2012). Practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* 61, 1000–1016. doi: 10.1016/j.neuroimage.2012.03.072

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.

Copyright © 2021 Oladosu, Liu, Pike, Koch, Metz 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 in Neuroscience

ORIGINAL RESEARCH
published: 13 May 2021
doi: 10.3389/fnins.2021.677106

Multi-Source Co-adaptation for EEG-Based Emotion Recognition by Mining Correlation Information

Jianwen Tao†* and *Yufang Dan*†

Institute of Artificial Intelligence Application, Ningbo Polytechnic, Zhejiang, China

Since each individual subject may present completely different encephalogram (EEG) patterns with respect to other subjects, existing subject-independent emotion classifiers trained on data sampled from cross-subjects or cross-dataset generally fail to achieve sound accuracy. In this scenario, the domain adaptation technique could be employed to address this problem, which has recently got extensive attention due to its effectiveness on cross-distribution learning. Focusing on cross-subject or cross-dataset automated emotion recognition with EEG features, we propose in this article a robust multi-source co-adaptation framework by mining diverse correlation information (MACI) among domains and features with $l_{2,1}-norm$ as well as correlation metric regularization. Specifically, by minimizing the statistical and semantic distribution differences between source and target domains, multiple subject-invariant classifiers can be learned together in a joint framework, which can make MACI use relevant knowledge from multiple sources by exploiting the developed correlation metric function. Comprehensive experimental evidence on DEAP and SEED datasets verifies the better performance of MACI in EEG-based emotion recognition.

Keywords: electroencephalogram, emotion recognition, multi-source adaptation, feature selection, maximum mean discrepancy

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Hongjiang Wei,
Shanghai Jiao Tong University, China
Cai Zhenhua,
Wuhan University of Technology, China

***Correspondence:**
Jianwen Tao
2019025@nbpt.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *07 March 2021*
Accepted: *22 March 2021*
Published: *13 May 2021*

Citation:
Tao J and Dan Y (2021) Multi-Source Co-adaptation for EEG-Based Emotion Recognition by Mining Correlation Information. Front. Neurosci. 15:677106. doi: 10.3389/fnins.2021.677106

INTRODUCTION

Although emotion can be easily captured by human beings due to its close relationship with human's cognition (Dolan, 2002), it cannot be readily recognized by instruments due to its complexity. Recently, as one of the most active research topics from the affective computing community, affection recognition had obtained a large amount of attention from pattern recognition and machine vision research fields (Kim et al., 2013). Generally, there are two categories on the responses of human emotion, i.e., external and internal responses. In this work, we focus on the latter. Basically, the representative internal responses include blood pressure, heart rate, respiration rate, electroencephalography (EEG), magneto encephalogram (Mühl et al., 2014), etc. Usually, the core components of a traditional emotion recognition system based on EEG are feature extraction and emotion classification (Lan et al., 2018). Practically, the time domain, frequency domain, and time–frequency domain are the main sources of the extracted EEG features (Jenke et al., 2014; Zhang et al., 2020b). The EEG feature extraction methods are more

comprehensively reviewed in Jenke et al. (2014). In the past few years, aiming at the problem of emotion classification, a large number of emotion recognition models based on EEG signals have been proposed (Musha et al., 1997; Kim et al., 2013). For instance, a new group sparse canonical correlation analysis method was proposed in Zheng (2017) for simultaneous EEG channel selection and emotion recognition; in Li et al. (2018c), a graph regularized sparse linear regression method was proposed to deal with EEG-based emotion recognition. In the recent years, the deep learning method based on EEG has shown better performance than those traditional methods in emotion recognition and widely exploited in feature extraction and emotion recognition at the same time. For example, Zheng et al. (2015) employed deep belief network for EEG-based emotion recognition; Song et al. (2018) modeled the multi-channel EEG features by utilizing a graph and then performed EEG-based emotion recognition on these features; the work by Li et al. (2018b) had proposed a novel neural network model for EEG-based emotion recognition task.

While many models and methods for emotion recognition based on EEG have been proposed, most of them worked well only in the scenario that the training and test data were from the same distribution or domain. Under this hypothesis, the classifier trained on the source domain can directly predict the labels of the target data. However, for the problem of cross-domain emotion recognition based on EEG, many EEG-based emotion recognition methods would fail because of the distribution mismatch of EEG features. To this end, domain adaptation (DA) emotion recognition algorithms have emerged to investigate and address the automated emotion recognition problem (Chu et al., 2017), in which one has target domain with few or even none of labeled images by leveraging other related but different source/auxiliary domain(s) (Bruzzone and Marconcini, 2010). A typical example is the cross-subject EEG emotion recognition problem, in which the training and testing EEG data are from different subjects. To deal with the challenging cross-subject EEG emotion recognition problem, Pandey and Seeja (2019) proposed a subject-independent approach for EEG emotion recognition. Li et al. (2018a) proposed another method for cross-subject EEG emotion recognition. In the past decade, deep neural networks (Ganin et al., 2016; Li et al., 2018b) have also driven rapid progress in DA. The DA issues can be solved by the domain adversarial neural network (Ganin et al., 2016). It remains unclear, however, whether the performance of deep DA methods is really contributed by their deep feature representation, the fine-tuned classifiers, or is rather an outcome of the adaptation regularization terms (Ghifary et al., 2017).

Although the existing DA method has obvious effectiveness and efficiency in the special use of emotion recognition (Chu et al., 2017) in order to make use of the correlation knowledge among domains and features, there is little work to use the joint feature selection method and then carry out the multi-source adaptive domain recognition of cross-dataset. Besides this, during DA, most of the multi-source domain adaptation (MDA) methods (Yang et al., 2007; Tommasi et al., 2014) generally cope with the sources independently without considering the correlation information among the source domains, which may destroy the discriminant structure (either intrinsic or extrinsic) of multi-source domains. Last but not the least, for a MDA system, it is crucial for source weight determination during learning based on the correlation and quality of source domains. To the best of our knowledge, these characters are not feasible enough in extant MDA methods.

In order to solve the above-mentioned problems in existing MDA, we exploit the relevant information among sources and features to learn a multi-source emotion recognition model. We mainly adopt the strategy of digging the relationship between multi-source domains and one target domain (including feature and distribution) for promoting multi-source adaptive emotion recognition. We aim to progress beyond existing works that have partially addressed those issues by exploring to solve all the above-mentioned issues in a unified framework. Specifically, we develop a robust multi-source co-adaptation method for EEG-based emotion recognition by employing the correlation information (MACI) among features and sources *via* $l_{2,1}-norm$ (Nie et al., 2010a) and correlation metric regularization. Under this framework, the correlation metric function is developed to mine the invariant knowledge among multi-source domains, the $l_{2,1}-norm$ loss function aims to reduce the influence of outliers or noise, and row sparsity is designed to obtain the solution of sparse feature selection (Zhang et al., 2020b). We match distributions between each domain pair (including both target and multi-source domains) by minimizing the nonparametric maximum mean discrepancy (MMD) (Gretton et al., 2009; Pan et al., 2011) in a reproducing kernel Hilbert space (RKHS). The contributions of this article are listed as follows:

1. We propose a unified multi-source adaptive emotion recognition framework with EEG features by combining $l_{2,1}-norm$ and correlation metric regularization.
2. Our framework selects features in a collaborative way and considers the correlated knowledge among features; the importance of each feature does not need to be evaluated separately. In addition, in our unified framework, we can learn multiple functions of feature selection for all source adaptation subjects synchronously so that our framework can use the correlated information of multiple sources as auxiliary information.
3. In this framework, the original geometric structure is retained by using the graph Laplacian regularization, and the $l_{2,1}-norm$ minimization sparse regression approach is used to suppress the influence of noise or outliers in the domains, which shows the robustness of the framework.
4. Through a large number of experiments on two EEG datasets, we prove the effectiveness and convergence of this framework.

The remainder of the paper is organized as follows. In section "Related Work," we discussed the related works with feature selection and MDA learning. In section "Proposed Framework," our framework MACI will be designed, while section "Algorithm" arranges the corresponding optimal algorithm of MACI. Section "Algorithm Analysis" gives algorithm analysis, including the convergence and generalization.

The experimental results and analysis on two real EEG datasets are presented in section "Experiments." Finally, we conclude in section "Conclusion."

RELATED WORK

In this section, we briefly review the prior emotion recognition with respect to EEG and multi-source adaptation techniques due to their relationships with our main ideas.

Multi-Source Domain Adaptation

In the past several years, the mismatch problem between source and target domains has been solved by many DA technologies, which are widely used in a large number of visual applications, such as image annotation/classification, video concept detection, target recognition, and so on (Yang et al., 2007; Duan et al., 2012b,c; Tao et al., 2015, 2016, 2017; Ghifary et al., 2017). In the existing works for conquering DA, discovering one or multiple domain-invariant classifier(s) is a widely research topic (Yang et al., 2007; Tommasi et al., 2014) by constructing certain common subspace to make different sources have the same (or similar) marginal distributions. Therefore, the source classifier would obtain well performance on the target domain. Several methods have been studied to measure the distribution similarities (Gretton et al., 2009), and the analysis from them (Mansour et al., 2009) show that the performance of the classifier on the target is in positive correlation with those similar sources.

Very recently, to overcome the so-called negative transfer issue (Rosenstein et al., 2005), MDA methodology has been put forward by leveraging knowledge from multiple sources (Zhang et al., 2015, 2019c). A common problem in MDA is how to reduce the distribution difference between domains (Ghifary et al., 2017). To solve this issue, existing MDA approaches can be simply grouped into two classes (Tao et al., 2019), i.e., classifier-centric learning and feature-centric learning. The former is mainly based on the learning of the source domain classifiers in the target domain to adjust for realizing the implicit adaptation in the target distribution (e.g., Yang et al., 2007; Duan et al., 2012c; Tao et al., 2012; Tommasi et al., 2014), while the latter tries to accomplish the distribution alignment by learning a new representation of the data through a certain transformation (e.g., Wang and Mahadevan, 2011; Ghifary et al., 2017). This article focuses on the research of unsupervised classifier approaches.

In real application scenarios, the classifier-centric MDA scheme usually aims to directly design multiple adaptive source classifiers by merging the multiple distributions' adaptation *via* feature representation or classifiers with model regularization. Lately, visual recognition works (Mansour et al., 2009; Nie et al., 2010a; Pan et al., 2011; Duan et al., 2012c; Tao et al., 2015, 2017) have proposed a great deal of classifier-centric MDA approaches. One part of classifier-centric MDA research assumes that there are enough number of unlabeled target instances and a large amount of labeled source instances in the training stage. Nevertheless, the remaining part of classifier-centric MDA research holds another hypothesis that only some labeled target instances are accessible in the training stage, which is also called model adaptation in the literature (Yang et al., 2007; Duan et al., 2012b,c; Tao et al., 2015, 2016, 2017; Ghifary et al., 2017). The model adaptation works effectively and efficiently just by exploiting the existing source models pre-trained on relevant but different source domains. Several representative state-of-the-arts include adaptive support vector machines (A-SVM) (Yang et al., 2007) *via* leveraging multiple source classifiers to suit a major target classifier, DA machine (or FastDAM) (Duan et al., 2012c) by employing sparsity regularizations and Laplacian manifold in least squares SVMs (Chai et al., 2016), etc. Recently, we also proposed some different model adaptation strategies (Tao et al., 2015, 2016, 2017) by leveraging the advantages of the low-rank and sparse representation.

Emotion Recognition

In recent research about affective computing, increasing attentions have been paid on emotion recognition in the community of brain–computer interfaces (BCIs) (Mühl et al., 2014; Chu et al., 2017). An ideal emotion-based BCI can detect the emotional state through spontaneous EEG signals without explicit input from the user (Zhang et al., 2019b) and make a corresponding response to different emotional states. This kind of BCI may enhance the consumer experience in the time of an interactive session. Therefore, different approaches in Zhang et al. (2016, 2017) have been designed to recognize various emotion signals from brain wave. The latest affective BCIs (aBCIs) took machine learning algorithms and depended on a few features with discriminative information (Jenke et al., 2014; Mühl et al., 2014). When recording EEG signals in order to generate a desired target emotion signal, it is necessary to provide users with affective stimulation of specific emotions. In the training/calibration session, the required features and corresponding emotion labels are extracted from EEG signals to train the classifier. In an ongoing BCI session, the feature extractor receives the real-time EEG data and then sends the extracted features to the classifier for real-time affection classification. In this paradigm (Mühl et al., 2014), many researchers have reported pleasing classification performance. However, even if the experimental results are encouraging, the performance of aBCI still could be impacted by some reason. Since the EEG-based emotion signals are different from subject to subject, it is indispensable to train a specific object classifier for the subject of interest. Even in the same subject, the EEG signals are unstable, and the earlier trained classifier may perform poorly in the same subject at a later time. Therefore, in order to maintain a satisfactory classification accuracy, it is necessary to recalibrate frequently.

Domain adaptation method (Judy et al., 2017; Tzeng et al., 2017; Ding et al., 2018) has nearly completely dominated the recent literature of BCI (Jayaram et al., 2016). In aBCI studies, Dolan (2002), Koelstra et al. (2012), Shi et al. (2013), Mühl et al. (2014), Zheng et al. (2015), Zheng and Lu (2015, 2016), Chai et al. (2016, 2017), Lan et al. (2018), Zhong et al. (2020), and search various DA approaches by exploiting the SEED dataset.

PROPOSED FRAMEWORK

Notations and Definitions

We describe the column vectors and matrices according to the small and capital letters, respectively, in this article. The often utilized symbols are listed in **Table 1**. The concatenation representation of k matrices according to row (horizontally) is like $[A_1, A_2, ..., A_k]$, and these matrices concatenation operations along a column (vertically) is denoted as $[A_1; A_2; ...; A_k]$. The $l_{2,1}-norm$ of A is defined as $||A||_{2,1} = \sum_{i=1}^{n} \left|\left|A_{i,:}\right|\right|_2 = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{d} A_{ij}^2}$, and the trace-norm of A is indicated as $||A||_* = tr((AA^T)^{\frac{1}{2}})$ (Lotfi and Akbarzadeh, 2014).

We mainly focus on the unsupervised MDA based on S various sources of c-class. Suppose there are n_a $(a = 1, 2, ..., S)$ instances in every source domain, respectively. In the $a-$th source domain, given $X^a = \{x_1^a, ..., x_{n_a}^a\} \in \mathrm{R}^{d \times n_a} \in \chi$, and it is a training instances matrix with c sub-classes, which are associated with their class labels $Y^a = [y_1^a, ..., y_{n_a}^a]^T \in \mathrm{R}^{n_a \times c} \in \Gamma = \{0, 1\}^{c \times 1}$, a target domain dataset is denoted as $X^t = \{x_1^t, x_2^t, ..., x_m^t\} \in \mathrm{R}^{d \times n_t} \in \chi$, with their pseudo-class labels $Y^t = [y_1^t, ..., y_{n_t}^t]^T \in \mathrm{R}^{n_t \times c} \in \Gamma$ obtained from some supervised models (e.g., SVMs) which are trained on the source domain with labeled data. Our ultimate goal is to recognize the ground-truth class of test data $x_k^t \in X^t$, under the conditions that each domain pair X^a and X^t is assumed to be of different marginal and conditional distributions. While we do not need to limit that the instances number in each source domain is identical with that assumed when shaped into the training matrix, for the sake of simplicity, we can extract the same number of training instances from each source domain.

We further denote by $X^{a(\bar{l})} \left(\bar{l} = 1, ..., c\right)$ the set of samples in X^a with the label $\bar{l}$. Similarly, the sample set in the target domain X^t with the label $\bar{l}$ is defined as $X^{t(\bar{l})}$. Note that the true labels of the set $X^{t(\bar{l})}$ are unknown. We therefore employ in this work a base classifier, e.g., SVM, to attribute pseudo-labels for the subset in the target domain. For easy expression, we further define the matrix $X_a = [X^a, X^t] \in \mathrm{R}^{d \times N}$ $(N = n_t + n_a)$ with its label matrix $Y_a = [Y^a, Y^t]$ in packing both source and target data with respect to the $a-$th source domain.

Definition 1 (MDA): Let $\Delta = \left\{\mathrm{P}^1, \ldots, \mathrm{P}^S\right\}$ be a set of S source domains and $\mathrm{P}^t \notin \Delta$ be a target domain. Denote by $_{X^a=\{x_i^a, y_i^a\}_{i=1}^{n_a} \sim \mathrm{P}^a (a=1,...,S)}$ the samples drawn from the $a-$th source domains and by $X^t = \left\{x_i^t\right\}_{i=1}^{n_t} \sim \mathrm{P}^t$ the samples drawn from the target domain. The task of MDA is to learn an ensemble function $f_{\mathrm{P}^t} : \mathcal{X} \to \Gamma$ by co-learning multiple classifiers given X^a $(a = 1, \ldots, S)$ and X^t as the training examples.

Definition 2 (Multi-source Domain Generalization): In this scenario, the target domain is inaccessible in the training stage. Given S source domains $\Delta = \left\{\mathrm{P}^1, \ldots, \mathrm{P}^S\right\}$ and denoted by $X^a = \left\{x_i^a, y_i^a\right\}_{i=1}^{n_a} \sim \mathrm{P}^a$ the samples drawn from the $a-$th source, the task of multi-source generalization is to co-learn multiple adaptive functions $f_{\mathrm{P}^a} : \mathcal{X} \to \Gamma$ only given $X^a, \forall a = 1, \ldots, S$ as the training examples, which could be well generalized to a certain unseen target domain.

TABLE 1 | Notations and descriptions.

Notations	Descriptions
n	Data size
d	Feature dimensionality of data
χ	Data space
Γ	Label space
$a = [a_1, a_2, ..., a_d]^T \in R^d$	Feature vector
$A \in R^{n \times d}$	Data matrix
$A_{i,j}$	The (i, j) entry of A
$A_{i,:}$ and $A_{:,j}$	The i-th row and j-th column of A
A^T and a^T	The transpose of matrix A and vector a
$tr(A)$	The trace of a matrix A
$\langle A_1, A_2 \rangle = tr(A_1^T A_2)$	The inner product of two matrices A_1 and A_2
$\|\|a\|\|_p := \left(\sum_{i=1}^{d} \|\|a_i\|\|^p\right)^{1/p}$	The p-norm of a vector a
$\|\|A\|\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{d} A_{i,j}^2}$	The Frobenius norm of A
I_r	Identity matrix of size $r \times r$
1_d	d-dimensional vector of ones
0_d	d-dimensional vector of zeroes

Problem Statement

In representative MDA, one can use the strategy of acquiring knowledge from multiple auxiliary sources to promote the target task of interest, which is better than learning each source task alone in emotion recognition. That is to say that common knowledge shared by multi-source domains is beneficial to emotion analysis. Moreover, some optimal recognition models have been developed in the latest works for the source domain and/or target domain separately. Furthermore, in these methods, joint multi-source adaption emotion recognition and feature selection has been largely unaddressed, and little or limited efforts have yet been devoted to the utilization of the correlated knowledge among sources.

To solve the above-mentioned issues, we propose in this work a robust multiple-source adaption emotion recognition method based on EEG features. The method utilizes the correlated knowledge among domains and features by joint $l_{2,1}-norm$ and correlation metric regularization and can process high-dimensional, sparse, outliers, and non-i.i.d EEG data at the same time. The designed method has three characteristics, which are integrated into a unified optimization formulation to find an effective emotion recognition model and align the feature distribution between source and target domains: (1) *via* employing the $l_{2,1}-norm$ minimization, a robust loss term is introduced to avoid the influence of noise or outliers in EEG signal, and a sparse regularization term is designed to eliminate over-fitting and a sparse feature subset is selected; (2) based on the designed regression model and the semantic distribution matching between each pair of domains, it not merely provides robustness on loss function but also retains the domain distribution (including local and global) structures and meanwhile maintains a high dependence on the (pseudo)-label knowledge of the source domains and the target domain (Zhang et al., 2020a) so as to obtain preferable generalization

performance; and (3) through our constructed metric function of correlation, we can make full use of the correlative information among multiple sources and transfer more discriminative knowledge to the target domain. To implement these properties, in the following part, we will detail the objective formulation of the proposed method.

General Formulation

In this section, we propose the general formulation of MACI framework underpinned by the robust regression principle and the regularization theory. In particular, our main purpose is to optimize a unified objective function by compromising the following three complementary objectives:

1. Robust multi-source co-regression with feature selection using $l_{2,1}-norm$ minimization, in which the domain label consistency is explicitly maximized through iterative linear label regression.
2. Aligning domain distributions including global statistical distributions and intra-domain semantic distributions or class conditional distributions.
3. Effectively utilizing correlation information among source domains *via* developing an effective correlation metric function.

For the multi-source adaptation emotion recognition of interest, we define the $a-$th $(a = 1, ..., S)$ classifier function as $f_a(X_a) = X_a^T W_a$, where W_a is the $a-$thclassifier model, and W_0 is certain reference model. Suppose there is a kernel feature map $\phi_a : \chi \rightarrow H_a$ [1] that projects the training data from the original feature space into certain RKHS (Nie et al., 2010b) H_a, the predictor weight W_a can be kernelized. We denote the kernel matrix as $(K_a)_{i,j} = \left\langle \phi(x_i^a), \phi(x_j^a) \right\rangle$, where $x_i^a, x_j^a \in X_a$. We present the empirical kernel map as discussed in Gretton et al. (2009):

$$\begin{aligned} &\phi_a : \quad \chi \rightarrow \mathrm{R}^N, \quad \text{for linear kernel mapping} \\ &x \rightarrow K_a(\cdot, x^a)|_{x_1^a, x_2^a, ..., x_N^a} = \left(K_a(x_1^a, x^a), ..., K_a(x_N^a, x^a)\right), \\ &\qquad\qquad \text{for nonlinear kernel mapping} \end{aligned}$$

We therefore have kernel matrices $K_a = \phi_a(X_a)$. Hence, the kernelized decision function on X_a $(a = 1, ..., S)$ becomes $f_a(X_a) = K_a^T W_a$. We further denote by $W = [W_1; ...; W_S]$ the concatenation matrix.

We then endeavor to find S cross-domain models parameterized by $\{W_a\}_{a=1}^S$ in some empirical RKHSs *via* jointly utilizing correlated knowledge among sources and features. In view of the above-cited objectives, we propose the following general formulation of MACI.

$$\Theta(W_a, F_a) = R(K_a^T W_a, Y_a) + \Omega_A(X^a, X^t) + Cor(W), \quad (1)$$

where $R(\cdot, \cdot)$ is the robust regression function with feature selection *via* $l_{2,1}-norm$ minimization, $\Omega_A(X^a, X^t)$ is certain distance metric function for aligning the domain distributions, and $Cor(\cdot)$ is a correlation metric function which is a global regularization term. In the subsequent sections, we focus on designing these components in the general formulation one by one to construct a unified framework.

[1] It is important to note that the feature mapping function ϕ_a $(1 \leq a \leq S)$ with respect to each source domain can be completely different from each other.

Design of Robust Multi-Source Co-regression With Feature Selection

To achieve the first objective mentioned above, one should jointly minimize each source regression loss and implement feature selection, in which the domain label consistency is explicitly maximized, and the data outliers are accounted for to avoid negative transfer. To this end, we first explain a predicted label matrix $F_a \in \mathrm{R}^{N \times c}$ $(a = 1, ..., S)$ into our predictive function (Nie et al., 2010b). The predicted values in this label matrix should satisfy local smoothness and global consistency, i.e., they should preserve the local geometry while fitting in with the true labels (Zhang et al., 2020a). To satisfy these requirements, we present a smooth regularization term on the label geometric structure between each source instance (Nie et al., 2010b; Yan et al., 2006), which is formulated as

$$g(F_a) = \left\{ \begin{array}{l} tr\left[\left(F_a - Y_a\right)^T \mathring{\mathrm{U}} \left(F_a - Y_a\right) \right] + \alpha tr(F_a L_a (F_a)^T) \\ \left|\left| F_a^T F_a = I_c, (F_a)_{i,j} \geq 0 \right.\right. \end{array} \right\},$$

where $\mathring{\mathrm{U}}$ is a diagonal matrix with $\mathring{\mathrm{U}}_{i,i} = \varsigma$ (ς is a large specified value) if $x_i^a \in X_a$ has a label, $\mathring{\mathrm{U}}_{i,i} = 0$ or else and a is a regularization parameter. L_a is the graph Laplacian matrix of the $a-$th source dataset, which is defined as $L_a = \Lambda_a - \prod_a$, where Λ_a is a diagonal matrix with $(\Lambda_a)_{i,i} = \sum_j (\prod_a)_{i,j}$, and $\prod_a$ is the weight matrix of the graph, which can be defined as:

$$(\prod_a)_{i,j} = \left\{ \begin{array}{c} \exp(-\gamma_a ||x_i^a - x_j^a||^2), \quad \text{if } x_i^a \in \delta_k(x_j^a) \text{ or } x_j^a \in \delta_k(x_i^a) \text{ and both have the same labels} \\ \exp(-\frac{\gamma_a}{||x_i^a - x_j^a||^2}), \quad \text{if } x_i^a \in \delta_k(x_j^a) \text{ or } x_j^a \in \delta_k(x_i^a) \text{ and both have different labels} \\ 0, \quad \text{otherwise} \end{array} \right.,$$

where the $_k$ nearest neighbors of x are assigned to $\delta_k(x)$, and γ_a is a hyper-parameter, which can be empirically selected as $\bar{\theta}_a \sqrt{c}$ by considering the impact of multi-class distribution on the affinity relationship among the domain data, where $\bar{\theta}_a$ is the square root of the mean norm of $_{X_a}$.

We therefore design the following multi-source sparse co-regression model for meeting the first objective.

$$\begin{aligned} &R(W_a^T \phi(X^a), F_a) = \sum_{a=1}^S \vartheta_a^{q_1} \left(||K_a^T W_a - F_a||_{2,1} + g(F_a) + \beta ||W_a||_{2,1} \right) \\ &s.t. \ \sum_{a=1}^S \vartheta_a = 1, \end{aligned} \quad (2)$$

where $\vartheta = [\vartheta_1, ..., \vartheta_a]^T$ is the weight vector to jointly combine all source regression loss, β is a regularization parameter, and $q_1 > 1$ is a tunable parameter for avoiding trivial solution. The model (2) is convex, and the $l_{2,1}-norm$ loss function $||K_a^T W_a - F_a||_{2,1}$ is robust to outliers (Li et al., 2015). In the meantime, the term $||W_a||_{2,1}$ assures that W_a can accomplish feature selection across different domains due to its sparsity. That is, by exploiting the correlation among different features, our approach can jointly evaluate all feature knowledge of source domains and target domain.

Design of Domain Distribution Alignment

As a nonparametric distribution discrepancy estimator, MMD (Gretton et al., 2009) was used to compare two distributions by transforming the distributions into a RKHS (Pan et al., 2011; Duan et al., 2012b; Tao et al., 2012; Chen et al., 2013; Long et al., 2014). Let $\mathcal{F}$ be a set of functions $f : \mathcal{X} \to \mathrm{R}$. The MMD between two domains P and Q is defined as

$$MMD_{\mathcal{F}}[\mathrm{P}, \mathrm{Q}] := \sup_{f \in \mathcal{F}} \left(\mathop{\mathrm{E}}_{\mathrm{P}}[f(x)] - \mathop{\mathrm{E}}_{\mathrm{Q}}[f(x)] \right). \quad (3)$$

The MMD measures the similarity level between two domains from the side of function class $\mathcal{F}$. To make the MMD a proper regularization for the classifier model W_a, we adopt the following the empirical estimation of MMD between X^a and X^t, which is defined as

$$MMD_e(X^a, X^t) := \left\| \frac{1}{n_a} \sum_{i=1}^{n_a} f_a(x_i^a) - \frac{1}{n_t} \sum_{j=1}^{n_t} f_a(x_j^a) \right\|_H^2 = tr(W_a^T K_a M_a K_a W_a), \quad (4)$$

where $||\cdot||_H$ is the RKHS norm, $(K_a)_{i,j} = < \phi_a(x_i^a), \phi_a(x_j^a) >$ with $x_i^a, x_j^a \in X_a$, and

$$M_{i,j}^a = \begin{cases} \frac{1}{n_a^2}, \text{ when } x_i^a, x_j^a \in X^a \\ \frac{1}{n_t^2}, \text{ when } x_i^a, x_j^a \in X^t \\ \frac{-1}{n_a n_t}, \text{ otherwise} \end{cases}. \quad (5)$$

As for $MMD_e(X^a, X^t)$ in Eq. 4, whereas even if there is a perfect domain distribution match, it does not assure the instances from different domains, but the same class of labels will be mapped near the transform space. Lack of semantic consistency will be a major reason for performance degradation. Therefore, we use the following terms to develop a semantically matched MMD (Long et al., 2014):

$$\begin{aligned} &MMD_{CA}(X^a, X^t) = \sum_{\bar{l}=1}^{c} MMD_{\varepsilon}(X^{a(\bar{l})}, X^{t(\bar{l})}) \\ &= \sum_{\bar{l}=1}^{c} tr(W_a^T K_{a(\bar{l})} M_{a(\bar{l})} K_{a(\bar{l})} W_a), \end{aligned} \quad (6)$$

where $K_{a(l)} = \phi_a([X^{a(\bar{l})}, X^{t(\bar{l})}])$ with $(K_{a(l)})_{i,j} = < \phi_a(x_i^{a(\bar{l})}), \phi_a(x_j^{a(\bar{l})}) >$, $x_i^{a(\bar{l})} \in X^{a(\bar{l})}$ and $x_j^{t(\bar{l})} \in X^{t(\bar{l})}$, and

$$(M_{a(\bar{l})})_{i,j} = \begin{cases} \frac{1}{m_{a\bar{l}}^2}, \text{ when } x_i^{a(\bar{l})}, x_j^{a(\bar{l})} \in X^{a(\bar{l})} \\ \frac{1}{m_{\bar{l}}^2}, \text{ when } x_i^{a(\bar{l})}, x_j^{a(\bar{l})} \in X^{t(\bar{l})} \\ \frac{-1}{n_{a\bar{l}} m_{\bar{l}}}, \text{ otherwise} \end{cases}. \quad (7)$$

We call Eq. 7 conditional (or semantic) MMD, which explicitly encourages instances from various domains but with the same label to map to the nearest in multi-source subspace. Finally, we suggest that the domain distribution alignment could be approached by learning multiple optimal models such that

$$\begin{aligned} &\Omega_A(X^a, X^t) = MMD_e(X^a, X^t) + MMD_{CA}(X^a, X^t) \\ &= tr(W_a^T K_a M_a K_a W_a) + \sum_{\bar{l}=1}^{c} tr(W_a^T K_{a(\bar{l})} M_{a(\bar{l})} K_{a(\bar{l})} W_a) \\ &= \sum_{\bar{l}=0}^{c} tr(W_a^T C_{a(\bar{l})} W_a) = tr(W_a^T \sum_{\bar{l}=0}^{c} C_{a(\bar{l})} W_a), \end{aligned} \quad (8)$$

where $C_{a(0)} = K_a M_a K_a$ and $C_{a(\bar{l})} = K_{a(\bar{l})} M_{a(\bar{l})} K_{a(\bar{l})}$ $\left(\bar{l} = 1, ..., c\right)$. Let $C_a = \sum_{\bar{l}=0}^{c} C_{a(\bar{l})}$, then we have $\Omega_A(X^a, X^t) = tr(W_a^T C_a W_a)$.

Design of Correlation Metric Function

As we know, a commonly used strategy in extant classifier-centric adaptation methods (Duan et al., 2012c; Tommasi et al., 2014) is to directly match the discriminant models between different domains, which is defined as:

Definition 3 (model discriminant discrepancy, MDD): Let $\mathcal{W}$ be a set of function parameters $\mathcal{W} : \mathcal{X} \to \mathrm{R}$. The model discriminant discrepancy between domains P and Q is defined as

$$MDD_{\mathcal{W}}[\mathrm{P}, \mathrm{Q}] := \sup_{W_{\mathrm{P}}, W_{\mathrm{Q}} \in \mathcal{W}} \left|\left| W_{\mathrm{P}} - W_{\mathrm{Q}} \right|\right|_F^2.$$

It may be difficult to push these two models respectively learnt from different domains when the distribution discrepancy between them is large. In our correlation metric function, we instead aim to guarantee each source model to be aligned with a global reference matrix W_0 so as to enable different source models to share the common knowledge for effectively utilizing correlation information among source domains. In essence, W_0 builds a transformation among source domains so that knowledge of one source can be used to another. They yield the following model alignment function (MAF):

Definition 4 (model alignment function): Given S domains $\{X^a\}_{a=1}^S$ on $\mathcal{X}$, we can think of the classification model set $\{W^a\}_{a=1}^S$ in some latent spaces. Their MAF is then defined as

$$\Psi\left(\{W_a\}_{a=1}^S\right) = \sum_{a=1}^{S} \eta_a ||W_a - W_0||_F^2$$

where $\eta = [\eta_1, ..., \eta_S]^T$ is a weight vector for discriminatively selecting different source knowledge with $\sum_{a=1}^S \eta_a = 1$, and W_0 is certain shared (common) discriminant model among these domains (Zhang et al., 2019a).

In essence, the MAF measures the similarity between two domain classifiers by the classification model. The next theorem is about from MAF to MDD between two domains.

Theorem 1 (MAF bounds MDD): The (squared) maximum discriminant discrepancy between domains P and Q on $\mathcal{X}$ is upper-bounded by their MAF with $\eta_{\mathrm{P}} = \eta_{\mathrm{Q}} = \eta_\Delta$:

$$\eta_\Delta MDD_{\mathcal{W}}^2[\mathrm{P}, \mathrm{Q}] \le \Psi\left(\left\{W_{\mathrm{P}}, W_{\mathrm{Q}}\right\}\right)$$

where $\mathcal{W} = \{W : \mathcal{X} \to \mathrm{R} | W \text{ is the classifier model}\}$ and $W_{\mathrm{P}}, W_{\mathrm{Q}} \in \mathcal{W}$. Specially if $\mathcal{W}$ is induced by a characteristic kernel on $\mathcal{X}$, then $\Psi\left(\left\{W_{\mathrm{P}}, W_{\mathrm{Q}}\right\}\right) = 0$ if and only if $\mathrm{P} = \mathrm{Q}$.

Proof. By definition 4 and triangle inequality theorem,

$$\begin{aligned} &\Psi\left(\left\{W_{\mathrm{P}}, W_{\mathrm{Q}}\right\}\right) \\ &= \eta_{\mathrm{P}} ||W_{\mathrm{P}} - W_0||_F^2 + \eta_{\mathrm{Q}} ||W_{\mathrm{P}} - W_0||_F^2 \\ &= \eta_\Delta \left(||W_{\mathrm{P}} - W_0||_F^2 + \left|\left| W_{\mathrm{Q}} - W_0 \right|\right|_F^2 \right) \\ &\ge \eta_\Delta \left|\left| (W_{\mathrm{P}} - W_0) - (W_{\mathrm{Q}} - W_0) \right|\right|_F^2 \\ &= \eta_\Delta \left|\left| W_{\mathrm{P}} - W_{\mathrm{Q}} \right|\right|_F^2 = \eta_\Delta MDD_{\mathcal{W}}^2[\mathrm{P}, \mathrm{Q}], \end{aligned}$$

that is, $\Psi\left(\{W_{\mathrm{P}}, W_{\mathrm{Q}}\}\right)$ bounds $MDD^2_{\mathcal{W}}[\mathrm{P}, \mathrm{Q}]$. If $\mathcal{W}$ is induced by some characteristic kernel on $\mathcal{X}$, then $\Psi\left(\{W_{\mathrm{P}}, W_{\mathrm{Q}}\}\right) = 0$ if and only if $\mathrm{P} = \mathrm{Q}$, which can be concluded from the result of Theorem 2.2 of Gretton et al. (2009).

Theorem 1 also indicates that the MAF is an effective metric method if the kernel on $\mathcal{X}$ is characteristic (Gretton et al., 2009).

Theorem 2: In particular, if $W_0 = \eta_{\mathrm{P}} W_{\mathrm{P}} + \eta_{\mathrm{Q}} W_{\mathrm{Q}}$ in MAF, we obtain

$$\Psi\left(\{W_{\mathrm{P}}, W_{\mathrm{Q}}\}\right) = \eta_{\mathrm{P}}\eta_{\mathrm{Q}} MDD^2_{\mathcal{W}}[\mathrm{P}, \mathrm{Q}] \leq \frac{1}{4} MDD^2_{\mathcal{W}}[\mathrm{P}, \mathrm{Q}].$$

Proof. By $\eta_P + \eta_Q = 1$, we have

$$\begin{aligned}
&\Psi\left(\{W_{\mathrm{P}}, W_{\mathrm{Q}}\}\right) \\
&\quad = \eta_P \left|\left| W_{\mathrm{P}} - \left(\eta_{\mathrm{P}} W_{\mathrm{P}} + \eta_{\mathrm{Q}} W_{\mathrm{Q}}\right)\right|\right|_F^2 \\
&\qquad + \eta_Q \left|\left| W_{\mathrm{Q}} - \left(\eta_{\mathrm{P}} W_{\mathrm{P}} + \eta_{\mathrm{Q}} W_{\mathrm{Q}}\right)\right|\right|_F^2 \\
&\quad = \eta_P \left|\left| \eta_{\mathrm{Q}} W_{\mathrm{P}} - \eta_{\mathrm{Q}} W_{\mathrm{Q}}\right|\right|_F^2 + \eta_Q \left|\left| \eta_{\mathrm{P}} W_{\mathrm{P}} - \eta_{\mathrm{P}} W_{\mathrm{Q}}\right|\right|_F^2 \\
&\quad = \left(\eta_{\mathrm{P}}\eta_{\mathrm{Q}}^2 + \eta_{\mathrm{Q}}\eta_{\mathrm{P}}^2\right) \left|\left| W_{\mathrm{P}} - W_{\mathrm{Q}}\right|\right|_F^2 = \eta_{\mathrm{P}}\eta_{\mathrm{Q}} MDD^2_{\mathcal{W}}[\mathrm{P}, \mathrm{Q}] \\
&\quad \leq \tfrac{1}{4} MDD^2_{\mathcal{W}}[\mathrm{P}, \mathrm{Q}],
\end{aligned}$$

where the last inequality follows after observing that $\eta_{\mathrm{P}}\eta_{\mathrm{Q}} \leq 1/4$ with the equality holding when $\eta_{\mathrm{P}} = \eta_{\mathrm{Q}} = 1/2$.

For achieving the third target of ours, the correlation metric function therefore was designed as follows:

$$Cor(W) = \sum_{a=1}^{S} \eta_a^{q_2} ||W_a - W_0||_F^2 + \frac{\lambda}{2} ||W||_*, \tag{9}$$

where $q_2 > 1$ is another tunable parameter to avoid a trivial solution. The regularization term $||W||_*$ in Eq. 9 enables different projection functions $\{W_a\}_{a=1}^S$ to share common information/parts through models of sources. Thus, the knowledge from multiple sources can be further shifted from one by one source domain.

Final Formulation

By using empirical kernel function and integrating Eqs 2, 8, and 9, we therefore propose the following unified framework to implement the combination of feature selection and domain adaptive learning by utilizing the correlated knowledge across multi-sources and features.

$$\begin{aligned}
&\min_{W_a, F_a, \vartheta_a, \eta_a} \sum_{a=1}^{S} \left(\vartheta_a^{q_1} \left(\left|\left|K_a^T W_a - F_a\right|\right|_{2,1} + tr(W_a^T C_a W_a)\right)\right. \\
&\qquad \left. + g(F_a) + \beta ||W_a||_{2,1}\right) + \sum_{a=1}^{S} \eta_a^{q_2} \left|\left| W_a - W_0\right|\right|_F^2 + \tfrac{\lambda}{2} ||W||_* \\
&s.t.\ \sum_{a=1}^{S} \vartheta_a = \sum_{a=1}^{S} \eta_a = 1.
\end{aligned} \tag{10}$$

Note that, in Eq. 10, the $l_{2,1} - norm$ loss function makes the model effectively robust to noises or outliers from domains. In addition, after $l_{2,1} - norm$ regularization is added to W_a, many rows in $W_a (a = 1, ..., S)$ become zero. Therefore, the characteristics corresponding to these rows with zeros are not significant for the target task learning. Therefore, for acquiring competitive performance, we can select features from the original domain data. Similarly (Nie et al., 2010a), we sort the rows in W_a by descending sequence in light of each row of values in $l_2 - norm$ and followed by selecting the top rows as the feature selection outcome.

Remark 1: Our proposed method (10) has some competitions in universalization and efficiency. Firstly, in order to deal with the change of feature dimensions and types in among source domains easily, we learn a separate classification model for each independent domain pair (i.e., one source domain and one target domain). Then, for dealing with more heterogeneous sources, the algorithm is easy to extend. Moreover, in order to improve the speed of the algorithm, the redundant and irrelevant knowledge in the original features is thrown away before classification through sparse feature selection method, and then the classification models are learned. Furthermore, joint $l_{2,1} - norm$ and trace-norm minimization can be used to express well the together learning of feature selection and classification models so as to assure that the common subspace in the multi-domain can be extracted.

ALGORITHM

In this section, we first give an iterative approach to optimize the objective function (10) followed by its complexity and classification function. Although matrix completion is realized by an alike optimization method from Yang et al. (2013), we focus on the other issue, that is, joint optimization of trace norm and $l_{2,1} - norm$. We then further present an effective and valuable extension to domain generalization when the target domain is inaccessible.

Algorithm Optimization

According to Nie et al. (2010a), the derivative of $tr\left(\mathring{\mathrm{T}}^T Q \mathring{\mathrm{T}}\right)$ is equal to the derivative of $\left|\left|\mathring{\mathrm{T}}\right|\right|_{2,1}$, i.e., $2tr\left(\mathring{\mathrm{T}}^T Q \mathring{\mathrm{T}}\right) = \left|\left|\mathring{\mathrm{T}}\right|\right|_{2,1}$, where Q is a diagonal matrix and its $i-$th diagonal value is $Q_{ii} = \frac{1}{2||\mathring{\mathrm{T}}_{i,:}||_2}$, and if $\mathring{\mathrm{T}}_{i,:} = 0$, we can let $Q_{ii} = \frac{1}{2||\mathring{\mathrm{T}}_{i,:}||_2 + \varepsilon}$, where ε is a very small given value. Hence, we can farther transform Eq. 10 into Eq. 11:

$$\begin{aligned}
&\min_{W_a, F_a, \vartheta_a, \eta_a} \sum_{a=1}^{S} (\vartheta_a^{q_1} tr(T_a^T Z_a T_a) + \vartheta_a^{q_1} tr(W_a^T C_a W_a) \\
&\qquad + g(F_a) + \beta tr(W_a^T G_a W_a)) \\
&\qquad + \sum_{a=1}^{S} \eta_a^{q_2} \left|\left| W_a - W_0\right|\right|_F^2 + \tfrac{\lambda}{2} tr\left(W^T \left(WW^T\right)^{-\frac{1}{2}} W\right) \\
&s.t.\ \sum_{a=1}^{S} \vartheta_a = \sum_{a=1}^{S} \eta_a = 1,
\end{aligned} \tag{11}$$

where $S_a = [\underbrace{0_r, ..., 0_r}_{a-1}, I_r, \underbrace{0_r, ..., 0_r}_{S-a}]^T$, the diagonal matrix G_a is based on W_a, where the $k-$th element is equal to $(G_a)_{kk} = \frac{1}{2||(W_a)_{k,:}||_2}$, the diagonal matrix Z_a is based on $T_a = \sqrt{\vartheta_a^{q_1}}(K_a^T W_a - F_a)$, and where the $k-$th entry is computed

by $(Z_a)_{kk} = \frac{1}{2||(T_a)_{k,:}||_2}$. By taking the derivative of Eq. (11) in reference to W_0 and equaling to zero, we obtain:

$$W_0 = W\eta_0, \tag{12}$$

where $\eta_0 = [\tilde{\eta}_1 I_r; ...; \tilde{\eta}_S I_r]$ with $\tilde{\eta}_a = \eta_a^{q_2} \Big/ \sum_{a=1}^{S} \eta_a^{q_2}$ $(a = 1, ..., S)$. Substituting Eq. 12 into Eq. 11, we have

$$\begin{aligned} \min_{W_a, F_a, \vartheta_a, \eta_a} & \sum_{a=1}^{S} \{tr\left(T_a^T Z_a T_a\right) + \vartheta_a^{q_1} tr(W_a^T C_a W_a) \\ & + g(F_a) + \beta tr\left(W_a^T G_a W_a\right)\} \\ & + \sum_{a=1}^{S} \eta_a^{q_2} ||WS_a - W\eta_0||_F^2 + \tfrac{\lambda}{2} tr\left(W^T \left(WW^T\right)^{-\frac{1}{2}} W\right). \end{aligned} \tag{13}$$

Note that the sub-gradient matrices Z_a and G_a in Eq. 13 are dependent on the matrices W_a and F_a, which are also unknown beforehand. Consequently, the objective function in Eq. 13 is a multi-variable optimization problem involving the variables W_a, F_a, ϑ_a, η_a. Since optimizing these variables simultaneously is difficult, we exploit the alternating iterative strategy to update one variable iteratively while the other variable(s) is(are) fixed. Therefore, the problem in Eq. 13 can be decomposed into four sets of convex sub-problems. We aim to find the optimal solution to each sub-problem alternatively and iteratively so that the objective function in Eq. 13 would converge to a local optimal solution. By initializing W_a and F_a, thus initializing Z_a and G_a, then we can start the iterations.

Optimize W_a and F_a by Fixing ϑ and η

By solving the derivative of Eq. 13 in reference to W_a and equaling to zero, we obtain:

$$W_a = \vartheta_a^{q_1} \left(\vartheta_a^{q_1} C_a + \vartheta_a^{q_1} K_a Z_a K_a + \beta G_a + \Omega + \lambda U\right)^{-1},$$

$$K_a Z_a F_a = E_a F_a \tag{14}$$

where

$$U = \frac{1}{2}\left(WW^T\right)^{-\frac{1}{2}}, \tag{15}$$

$$\Omega = \sum_{a=1}^{S} \left(\eta_a^{q_2} \left(S_a - \eta_0\right)^T \left(S_a - \eta_0\right)\right), \tag{16}$$

and

$$E_a = \vartheta_a^{q_1} \left(\vartheta_a^{q_1} C_a + \vartheta_a^{q_1} K_a Z_a K_a + \beta G_a + \Omega + \lambda U\right)^{-1} K_a Z_a. \tag{17}$$

Plugging Eq. 14 into Eq. 13, by mathematical calculating, we can get:

$$tr\left(F_a^T N_a F_a\right) + \alpha tr(F_a L_a (F_a)^T) + tr\left[\left(F_a - Y_a\right)^T \mathring{U} \left(F_a - Y_a\right)\right], \tag{18}$$

where

$$N_a = E_a^T \left(C_a + \beta G_a + \lambda U\right) E_a + \alpha L_a + \left(K_a^T E_a - I_n\right)^T Z_a \left(K_a^T E_a - I_n\right).$$

Lastly, substituting the optimal solution of the other variables into Eq. 18 to update F_a, the constraints $F_a^T F_a = I_c$ should be added additionally, and $(F_a)_{ij} \geq 0$. Then, we can get the objective function in reference to F_a :

$$\begin{aligned} \Theta(F_a) = \min_{F_a} & tr\left(F_a^T \left(N_a + \alpha L_a\right) F_a\right) + \\ & tr\left[\left(F_a - Y_a\right)^T \mathring{U} \left(F_a - Y_a\right)\right] + \frac{\zeta}{2}\left|\left|F_a^T F_a - I_c\right|\right|_F^2 + tr\left(\theta F_a^T\right), \end{aligned} \tag{19}$$

where ζ [2] and θ are balance arguments. By solving the derivative of Eq. 19 in reference to $F_{i,j}$ and equaling to zero and exploiting the K.K.T. with constraint term $\theta_{ij} F_{i,j} = 0$, we can get:

$$\begin{aligned} \tfrac{\partial \Theta(F_a)}{\partial F_a} &= \left(N_a + \alpha L_a + \mathring{U}\right) F_a - \mathring{U} Y_a + \zeta F_a \left(F_a^T F_a - I_c\right) \\ &\quad + \theta/2 = 0 \\ &\Rightarrow (F_a)_{i,j} \leftarrow (F_a)_{i,j} \tfrac{\left(\mathring{U} Y_a + \zeta F_a\right)_{i,j}}{\left(\zeta F_a F_a^T F_a + \left(N_a + \alpha L_a + \mathring{U}\right) F_a\right)_{i,j}}. \end{aligned} \tag{20}$$

Optimize ϑ_a by Fixing W_a, F_a, and η_a

In this situation, the issue in Eq. 13 changes to a small problem as follows:

$$\min_{\vartheta_a \geq 0, \vartheta_a^T 1 = 1} \sum_{a=1}^{S} \left\{\vartheta_a^{q_1} tr\left(T_a^T Z_a T_a\right) + \vartheta_a^{q_1} tr(W_a^T C_a W_a)\right\}. \tag{21}$$

Let $g_a = tr\left(T_a^T Z_a T_a\right) + tr(W_a^T C_a W_a)$, the Lagrange function of Eq. 21 is

$$\Im(\vartheta_a, \phi) = \sum_{a=1}^{S} \vartheta_a^{q_1} g_a - \phi\left(\sum_{a=1}^{S} \vartheta_a - 1\right). \tag{22}$$

Setting the derivative of $\Im(\vartheta_a, \phi)$ with respect to ϑ_a is equivalent to 0, and we can obtain:

$$\vartheta_a = \left(\phi \big/ (q_1 g_a)\right)^{\frac{1}{q_1 - 1}}. \tag{23}$$

Substituting Eq. 23 into the constraint $\sum_{a=1}^{S} \vartheta_a = 1$, we obtain

$$\vartheta_a = (g_a)^{1/(1-q_1)} \Big/ \sum_{a=1}^{S} (g_a)^{1/(1-q_1)}. \tag{24}$$

Optimize η_a by Fixing W_a, F_a, and ϑ_a

By fixing W_a, F_a, and ϑ_a, the problem in Eq. 13 then becomes the following sub-problem:

$$\min_{\eta_a \geq 0, \eta_a^T 1 = 1} \sum_{a=1}^{S} \eta_a^{q_2} ||WS_a - W\eta_0||_F^2. \tag{25}$$

[2]For maximal consistency between $F^T F$ and I_c, the parameter ζ should be set as a relatively large value. In our experiments, we therefore empirically set $\zeta = 10^3$ without loss of performance to some extent.

Let $h_a = ||WS_a - W\eta_0||_F^2$, the Lagrange function of Eq. 25 is

$$\Im(\eta_a, \rho) = \sum_{a=1}^{S} \eta_a^{q_2} h_a - \rho\left(\sum_{a=1}^{S} \eta_a - 1\right). \quad (26)$$

Setting the derivative of $\Im(\eta_a, \rho)$ in reference to η_a is equivalent to 0, and we then get:

$$\eta_a = \left(\rho \big/ (q_2 h_a)\right)^{\frac{1}{q_2 - 1}}. \quad (27)$$

Substituting Eq. 27 into the constraint $\sum_{a=1}^{S} \eta_a = 1$, we obtain:

$$\eta_a = (h_a)^{1/(1-q_2)} \Big/ \sum_{a=1}^{S} (h_a)^{1/(1-q_2)}. \quad (28)$$

Overall Procedure

In this sub-section, we finally report the whole optimization process of MACI in Algorithm 1, where a window-based breaking criterion is employed to better obtain the convergence of the algorithm (Zhang et al., 2019a). Concretely speaking, defining a window size $\hbar$, we compute $\varsigma = ||Max\Theta_{itr} - Min\Theta_{itr}|| \big/ Max\Theta_{itr}$ in $itr-$th iteration, where $\Theta_{itr} = \{Obj_{itr-\hbar+1}, ..., Obj_{itr}\}$ represents the set of historical target values in the window. When ς is less than a given threshold ε, that is $\varsigma < \varepsilon$, the algorithm stops iterating. In our experiments, we set $\varepsilon = 10^{-5}$ and $\hbar = 6$ empirically without losing statistical performance. We will discuss in section "Convergence" why Algorithm 1 is convergent.

Algorithm 1. Multiple sources adaptation learning by utilizing correlation knowledge.

Input: Source datasets $\{X_i^s\}_{i=1}^S$, $\{L_i\}_{i=1}^S$, target dataset X^t, and parameters α, β, and λ, the maximal iteration number ℓ.
Output: Converged projection matrices $\{W_i\}_{i=1}^S$, and matrices $\{F_i\}_{i=1}^S$ and W_0.
Initialization: Set $itr = 0$, and initialize $\{W_i^{itr}\}_{i=0}^S$ randomly. Let $W^{itr} = [W_1^{itr}, ..., W_S^{itr}]$;
1: for $i = 1$ to S do
{
Compute matrix M_i^{itr} and $M_{i(l)}^{itr}$, and K_i^{itr} and $K_{i(l)}^{itr}$ with empirical kernel mapping, thus computing $C_i^{itr} = \sum_{l=0}^{c} C_{i(l)}^{itr}$ by $C_{i(0)}^{itr} = K_i^{itr} M_i^{itr} K_i^{itr}$ and $C_{i(l)}^{itr} = K_{i(l)}^{itr} M_{i(l)}^{itr} K_{i(l)}^{itr}$, $l = 1, ..., c$;
Compute η_i^{itr} by Eq. 28, and then construct matrix η^{itr} and $\eta_0^{itr} = [\eta_1^{itr} I_r; ...; \eta_S^{itr} I_r]$;
Compute $F_i^{itr} = K_i^T W_i^{itr}$;
}
2: repeat
{
Compute W_0^{itr} by (12);
Compute the diagonal matrix U^{itr} by (15);
Compute the matrix $\mathbf{\Omega}^{itr}$ by (16);
set $i = 1$;
repeat
{
Compute G_i^{itr} with respect to W_i^{itr} ;
Compute Z_i^{itr} with respect to $K_i^T W_i^{itr} - F_i^{itr}$;
Compute $g_i^{itr} = tr\left(\left(K_i^T W_i^{itr} - F_i^{itr}\right)^T Z_i \left(K_i^T W_i^{itr} - F_i^{itr}\right)\right)$
$+tr((W_i^{itr})^T C_i^{itr} W_i^{itr})$;
Compute ϑ_i^{itr} according to Eq. 24, and then construct ϑ^{itr} ;
Compute the matrix E_i^{itr} by Eq. 17, and then N_i^{itr} by Eq. 19;
Compute F_i^{itr} according to Eq. 20;
Compute W_i^{itr} according to Eq. 14;
$i = i + 1$;
} until $i > S$
Update $W_i^{itr+1} = W_i^{itr}$ s.t. $i = 1, .., S$;
Update $F_i^{itr+1} = F_i^{itr}$ according to (20) s.t. $i = 1, .., S$;
Update ϑ_i^{itr+1} according to (24) s.t. $i = 1, .., S$;
Update η_i^{itr+1} according to (27) s.t. $i = 1, .., S$;
Update W_0^{itr+1} according to (12);
Let $itr = itr + 1$;
}until $itr > \ell$ or $\varsigma < 10^{-5}$
3: return $\{W_a\}_{a=0}^S$ and $\{F_i\}_{i=1}^S$.

Computational Complexity

In this subsection, we give a formal analysis about the computational complexity of several main components in Algorithm 1 using the big O notation. Firstly, the construction of the $k-NN$ graph and computing of the kernel matrix $\{K_i\}_{i=1}^S$, respectively, need computational cost $O(Sdn_a^2)$ and $O(SdN^2)$. Then, the optimization proceeds according to step by step iteratively. The cost for computing F_a is $O\left(3n_a^3 + n_a^2 c\right)$. After F_a was updated, computing W_a would cost $O(n_a c^2 + dc^2)$. In a word, the whole calculating cost is $O\left(\ell S\left(3n_a^3 + n_a^2 c + n_a c^2 + dc^2\right) + Sdn_a^2 + SdN^2\right)$. We assume that all Laplacian matrixes $\{L_i\}_{i=1}^S$ can be pre-calculated before the iterative optimization of Algorithm 1, and multi-kernel can be pre-calculated and put into memory before training. Thus, Algorithm 1 is effective and efficient computationally.

Target Classification

The datasets of the unlabeled samples from the target domain are defined as $K_u^t = \{(k_j^t)\}_{j=1}^{n_t}$. In the unsupervised DA learning case, one can predict a target class using the classification model W_0. Specifically, one may use $\arg\max_{1 \le j \le c} (\Gamma_u^t)_j$ to classify a test sample $k_u^t \in K_u^t$ into one of the c target classes, where $\Gamma_u^t = (W_0)^T k_u^t$.

In our multi-source adaptation framework, nevertheless, another voting method defined as "sum" can be deduced, that is, once $\{W_a^s\}_{a=1}^S$ are obtained, for a test data $k_u^t \in K_u^t$, we can learn its label vector Γ_u^t by minimizing the residue between Γ_u^t and the projected vector of each source model:

$$\min_{\Gamma_u^t} \sum_{a=1}^{S} \vartheta_a \left|\left|\left(k_u^t\right)^T W_a^s - \Gamma_u^t\right|\right|_2^2. \quad (29)$$

The result of Eq. 29 can be acquired according to the constraint term $\sum_{a=1}^{S} \vartheta_a = 1$:

$$\Gamma_u^t = \sum_{a=1}^{S} \vartheta_a \left(k_u^t\right)^T W_a^s. \quad (30)$$

Once Γ_u^t is computed by using Eq. 30, we then use $\arg\max_{1 \le j \le c} (\Gamma_u^t)_j$ to determine the class for this test data.

ALGORITHM ANALYSIS

Convergence

We start with the next two lemmas and then demonstrate that the alternant optimization process, namely, step 2 in Algorithm 1, in the optimization issue of the Eq. 10, the optimal solution of $\{W_i\}_{i=1}^S$ converges.

Lemma 1. (Nie et al., 2010a) There are any two values $V_1, V_2 \in \mathrm{R}^d$, and they are not equal to zero; we can get the inequality as:

$$||V_1||_2 - \frac{||V_1||_2^2}{2||V_2||_2} \leq ||V_2||_2 - \frac{||V_2||_2^2}{2||V_2||_2}. \tag{31}$$

Lemma 2. (Nie et al., 2010a) For any invertible matrices $\mathring{\mathrm{P}}$ and $\mathring{\mathrm{Q}}$, the following inequality holds:

$$\frac{1}{2}tr\left(\mathring{\mathrm{P}}\mathring{\mathrm{Q}}^{-\frac{1}{2}}\right) - tr\left(\mathring{\mathrm{P}}^{\frac{1}{2}}\right) \geq \frac{1}{2}tr\left(\mathring{\mathrm{Q}}\mathring{\mathrm{Q}}^{-\frac{1}{2}}\right) - tr\left(\mathring{\mathrm{Q}}^{\frac{1}{2}}\right). \tag{32}$$

Then, the iterative method designed in Algorithm 1 can converge to the optimal solution, which will be proved in the next theorem.

Theorem 3: In each iteration of Algorithm 1, the objective function of issue in Eq. 10 will be monotonically decreasing and finally will converge to the optimal solution of the issue.

Proof. For easy description, we define the updated W_i and F_i in the iteration τ as W_i^τ and F_i^τ $(i = 1, ..., S)$ separately. The updating from step 2 of Algorithm 1 is equivalent to the optimum of the next problem:

$$\begin{aligned}\min_{W_i,F_i,\vartheta_i,\eta_i} &\sum_{i=1}^S \left\{\vartheta_i^{q_1}\left[tr\left(\left(K_i^T W_i - F_i\right)^T Z_i\left(K_i^T W_i - F_i\right)\right)\right.\right.\\ &\left.\left.+tr(W_i^T C_i W_i)\right] + \alpha g(F_i) + \beta tr\left(W_i^T G_i W_i\right)\right\}\\ &+tr\left(W^T \Omega W\right) + \lambda tr\left(W^T U W\right)\end{aligned}$$

Following the expressions of Z_i, G_i, and U, then we can get:

$$\begin{aligned}&\sum_{i=1}^S \left\{\begin{array}{l} tr\left((W_i^{\tau+1})^T C_i W_i^{\tau+1}\right) + \alpha g(F_i^{\tau+1})\\ +\vartheta_i^{q_1}\sum_{j=1}^n \frac{||Z_i(j,:)^{\tau+1}||_2^2}{||Z_i(j,:)^\tau||_2} + \beta\sum_{j=1}^n \frac{||W_i(j,:)^{\tau+1}||_2^2}{||W_i(j,:)^\tau||_2}\end{array}\right\}\\ &\quad+tr\left((W^{\tau+1})^T \Omega^{\tau+1} W^{\tau+1}\right) + \lambda tr\left((W^{\tau+1})^T U^\tau W^{\tau+1}\right)\\ &\leq \sum_{i=1}^S \left\{\begin{array}{l} tr\left((W_i^\tau)^T C_i W_i^\tau\right) + \alpha g(F_i^\tau)\\ +\vartheta_i^{q_1}\sum_{j=1}^n \frac{||Z_i(j,:)^\tau||_2^2}{||Z_i(j,:)^\tau||_2} + \beta\sum_{j=1}^n \frac{||W_i(j,:)^\tau||_2^2}{||W_i(j,:)^\tau||_2}\end{array}\right\}\\ &\quad+tr\left((W^\tau)^T \Omega^\tau W^\tau\right) + \lambda tr\left((W^\tau)^T U^\tau W^\tau\right).\end{aligned} \tag{33}$$

We can have the next inequality by Lemma 1:

$$\begin{aligned}&\sum_{j=1}^n\left(\left|\left|(W_i)_{j,:}^{\tau+1}\right|\right|_2 - \frac{\left|\left|(W_i)_{j,:}^{\tau+1}\right|\right|_2^2}{2\left|\left|(W_i)_{j,:}^\tau\right|\right|_2}\right)\\ &\qquad\leq \sum_{j=1}^n\left(\left|\left|(W_i)_{j,:}^\tau\right|\right|_2 - \frac{\left|\left|(W_i)_{j,:}^\tau\right|\right|_2^2}{2\left|\left|(W_i)_{j,:}^\tau\right|\right|_2}\right).\end{aligned} \tag{34}$$

Therefore, we have

$$\begin{aligned}&\sum_{i=1}^S \left\{\begin{array}{l} tr\left((W_i^{\tau+1})^T C_i W_i^{\tau+1}\right) + \alpha g(F_i^{\tau+1})\\ +\vartheta_i^{q_1}\sum_{j=1}^n \left|\left|(Z_i)_{j,:}^{\tau+1}\right|\right|_2 + \beta\sum_{j=1}^n \left|\left|(W_i)_{j,:}^{\tau+1}\right|\right|_2\end{array}\right\}\\ &\quad+tr\left((W^{\tau+1})^T \Omega^{\tau+1} W^{\tau+1}\right) + \lambda tr\left((W^{\tau+1})^T U^{\tau+1} W^{\tau+1}\right)\\ &\leq \sum_{i=1}^S \left\{\begin{array}{l} tr\left((W_i^\tau)^T C_i W_i^\tau\right) + \alpha g(F_i^\tau)\\ +\vartheta_i^{q_1}\sum_{j=1}^n \left|\left|(Z_i)_{j,:}^\tau\right|\right|_2 + \beta\sum_{j=1}^n \left|\left|(W_i)_{j,:}^\tau\right|\right|_2\end{array}\right\}\\ &\quad+tr\left((W^\tau)^T \Omega^\tau W^\tau\right) + \lambda tr\left((W^\tau)^T U^\tau W^\tau\right).\end{aligned} \tag{35}$$

Eq. 35 can be further rewritten as:

$$\begin{aligned}&\sum_{i=1}^S\{tr\left((W_i^{\tau+1})^T C_i W_i^{\tau+1}\right) + \alpha g(F_i^{\tau+1}) + \vartheta_i^{q_1}\\ &\sum_{j=1}^n\left|\left|(Z_i)_{j,:}^{\tau+1}\right|\right|_2 + \beta\sum_{j=1}^n\left|\left|(W_i)_{j,:}^{\tau+1}\right|\right|_2\}\\ &+tr\left((W^{\tau+1})^T \Omega^{\tau+1} W^{\tau+1}\right) + \lambda tr\left(W^{\tau+1}(W^{\tau+1})^T U^{\tau+1}\right)\\ &-\tfrac{\lambda}{2}tr\left(\left(W^{\tau+1}(W^{\tau+1})^T\right)^{\frac{1}{2}}\right) + \tfrac{\lambda}{2}tr\left(\left(W^{\tau+1}(W^{\tau+1})^T\right)^{\frac{1}{2}}\right)\\ &\leq \sum_{i=1}^S\{tr\left((W_i^\tau)^T C_i W_i^l\right) + \alpha g(F_i^\tau) + \vartheta_i^{q_1}\\ &\sum_{j=1}^n\left|\left|(Z_i)_{j,:}^\tau\right|\right|_2 + \beta\sum_{j=1}^n\left|\left|(W_i)_{j,:}^\tau\right|\right|_2\}\\ &+tr\left((W^\tau)^T \Omega^\tau W^\tau\right) + \lambda tr\left(W^\tau(W^\tau)^T U^\tau\right)\\ &-\tfrac{\lambda}{2}tr\left(\left(W^\tau(W^\tau)^T\right)^{\frac{1}{2}}\right) + \tfrac{\lambda}{2}tr\left(\left(W^\tau(W^\tau)^T\right)^{\frac{1}{2}}\right).\end{aligned} \tag{36}$$

Noting that $U^l = \frac{1}{2}\left(W^\tau(W^\tau)^T\right)^{-\frac{1}{2}}$ and according to Lemma 2, we get

$$\begin{aligned}&\lambda tr\left(W^{\tau+1}(W^{\tau+1})^T U^{\tau+1}\right) - \lambda tr\left(\left(W^{\tau+1}(W^{\tau+1})^T\right)^{\frac{1}{2}}\right)\\ &\qquad\geq \lambda tr\left(W^\tau(W^\tau)^T U^\tau\right) - \lambda tr\left(\left(W^\tau(W^\tau)^T\right)^{\frac{1}{2}}\right).\end{aligned} \tag{37}$$

Subtracting Eq. (37) from Eq. 36, we have

$$\begin{aligned}&\sum_{i=1}^S\{tr\left((W_i^{\tau+1})^T C_i W_i^{\tau+1}\right) + \alpha g(F_i^{\tau+1}) + \vartheta_i^{q_1}\\ &\sum_{j=1}^n\left|\left|(Z_i)_{j,:}^{\tau+1}\right|\right|_2 + \beta\sum_{j=1}^n\left|\left|(W_i)_{j,:}^{\tau+1}\right|\right|_2\}\\ &+tr\left((W^{\tau+1})^T \Omega^{\tau+1} W^{\tau+1}\right) + \lambda tr\left(\left(W^{\tau+1}(W^{\tau+1})^T\right)^{\frac{1}{2}}\right)\\ &\leq \sum_{i=1}^S\{tr\left((W_i^\tau)^T C_i W_i^\tau\right) + \alpha g(F_i^\tau) + \vartheta_i^{q_1}\\ &\sum_{j=1}^n\left|\left|(Z_i)_{j,:}^\tau\right|\right|_2 + \beta\sum_{j=1}^n\left|\left|(W_i)_{j,:}^\tau\right|\right|_2\}\\ &+tr\left((W^\tau)^T \Omega^\tau W^\tau\right) + \lambda tr\left(\left(W^\tau(W^\tau)^T\right)^{\frac{1}{2}}\right).\end{aligned} \tag{38}$$

That is to say

$$\begin{aligned}&\sum_{i=1}^S\{tr\left((W_i^{\tau+1})^T C_i W_i^{\tau+1}\right) + \alpha g(F_i^{\tau+1}) + \vartheta_i^{q_1}\\ &\left|\left|K_i^T W_i^{\tau+1} - F_i^{\tau+1}\right|\right|_{2,1} + \beta\left|\left|W_i^{\tau+1}\right|\right|_{2,1}\}\\ &+tr\left((W^{\tau+1})^T \Omega^{\tau+1} W^{\tau+1}\right) + \lambda\left|\left|W^{\tau+1}\right|\right|_*\\ &\leq \sum_{i=1}^S\{tr\left((W_i^\tau)^T C_i W_i^\tau\right) + \alpha g(F_i^\tau) + \vartheta_i^{q_1}\\ &\left|\left|K_i^T W_i^\tau - F_i^\tau\right|\right|_{2,1} + \beta\left|\left|W_i^\tau\right|\right|_{2,1}\}\\ &+tr\left((W^\tau)^T \Omega^\tau W^\tau\right) + \lambda\left|\left|W^\tau\right|\right|_*.\end{aligned} \tag{39}$$

Hence, the theorem has been verified.

According to the optimization strategy of Algorithm 1, the objective function is monotonically decreasing in problem Eq. 10, so it is easiest to observe that the algorithm is convergent.

Generalization

In this part, we derive an empirical bound for our method that shows how both MAF and PDS control the generalization performance under the situation of the squared loss $loss(a, b) = (a - b)^2$. The main idea is to merge the domain scatter into the proven adaptive range for the distance difference (Ghifary et al., 2017).

Denote by $H{:=}\{h : \mathcal{X} \to \mathcal{Y}\}$ a hypothesis class of functions in the RKHS $\mathcal{H}$, where $\mathcal{X}$ is a compact set and $\mathcal{Y}$ is a label space. Given a loss function $loss(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \to \mathrm{R}_+$ and a domain distribution $\mathcal{D}$ over $\mathcal{X}$, we denote by $\mathcal{L}_{\mathcal{D}}\left(h, \mathring{\mathrm{h}}\right) = E_{x\sim\mathcal{D}}\left[loss\left(h(x), \mathring{\mathrm{h}}(x)\right)\right]$ the expected loss for the given two functions $h, \mathring{\mathrm{h}} \in H$. Then, the distance of domain difference between two distributions P and Q is defined as:

$$disc(\mathrm{P}, \mathrm{Q}) = \sup_{h, \mathring{\mathrm{h}} \in H} \{\mathcal{L}_{\mathrm{P}}\left(h, \mathring{\mathrm{h}}\right) - \mathcal{L}_{\mathrm{Q}}\left(h, \mathring{\mathrm{h}}\right)\}, \tag{40}$$

By the notation in Eq. 40, we can obtain domain generalization bounds by domain scatter. Let f_{P} and f_{Q} be the true labeling functions for domain P and Q, respectively, and $h_{\mathrm{P}}^* := \underset{h\in H}{\operatorname{argmin}} \mathcal{L}_{\mathrm{P}}\left(h, f_{\mathrm{P}}\right)$ and $h_{\mathrm{Q}}^* := \underset{h\in H}{\operatorname{argmin}} \mathcal{L}_{\mathrm{Q}}\left(h, f_{\mathrm{Q}}\right)$ be the minimizers. The following theorem provides adaptation bounds with PDS (. . ..).

Theorem 4 (adaptation bounds with PDS) (Ghifary et al., 2017): Denote by $H{:=}\{f \in \mathcal{H} : \mathcal{X} \to \mathrm{R}, \left||f|\right|_{\mathcal{H}} \leq 1$ *and* $\left||f|\right|_{\infty} \leq r\}$ a class of functions in the RKHS $\mathcal{H}$ and by $X_{\mathcal{X}}^{\mathrm{P}} = \left(x_1^s, ..., x_{n_s}^s\right) \sim \mathrm{P}$ and $X_{\mathcal{X}}^{\mathrm{Q}} = \left(x_1^t, ..., x_{n_t}^t\right) \sim \mathrm{Q}$ the source and target dataset, respectively. Let $loss(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \to [0, \Upsilon]$ be a $q-$Lipschitz loss function, i.e., for all $a, b \in \mathcal{Y} \times \mathcal{Y}$, $\left|\left|loss(a) - loss(b)\right|\right| = q\left||a - b|\right|$. Then, for any hypothesis $h \in H$, with probability of at least $1 - \delta$, the following generalization bound holds with the Rademacher complexity $\Re_{X_{\mathcal{X}}^{\mathrm{P}}}(H)$ over $X_{\mathcal{X}}^{\mathrm{P}}$:

$$\begin{aligned}&\mathcal{L}_{\mathrm{Q}}\left(h, f_{\mathrm{Q}}\right) - \mathcal{L}_{\mathrm{Q}}\left(h_{\mathrm{Q}}^*, f_{\mathrm{Q}}\right) \leq \mathcal{L}_{\mathrm{P}}\left(h, h_{\mathrm{P}}^*\right) + 2q\Re_{X_{\mathcal{X}}^{\mathrm{P}}}(H)\\ &+3\Upsilon\sqrt{\frac{\log\frac{2}{\delta}}{2n_t}} + 8r\sqrt{\Xi_{\phi}(\{\mu_{\mathrm{Q}}, \mu_{\mathrm{P}}\})} + \mathcal{L}_{\mathrm{P}}\left(h_{\mathrm{Q}}^*, h_{\mathrm{P}}^*\right).\end{aligned} \tag{41}$$

Theorem 4 provides a generalization bound for DA by introducing PDS and Rademacher complexity that measures the level to which a class of functions can fit random noise. The Rademacher complexity measure is the basis of relating empirical loss with expected loss. From Theorem 4, for a successful DA, we shall make $\mathcal{L}_{\mathrm{P}}\left(h_{\mathrm{P}}^*, h_{\mathrm{Q}}^*\right)$ as small as possible. According to definition 3, the (squared) loss $\mathcal{L}_{\mathrm{P}}\left(h_{\mathrm{P}}^*, h_{\mathrm{Q}}^*\right)$ is essentially equivalent to MDD in some optimal RKHS. We then further provide the following adaptation bounds with PDS and MAF, which follows by Theorem 1 combined with Theorem 4.

Theorem 5 (adaptation bounds with PDS and MAF): Denote by $H{:=}\{f \in \mathcal{H} : \mathcal{X} \to \mathrm{R}, \left||f|\right|_{\mathcal{H}} \leq 1$ and $\left||f|\right|_{\infty} \leq r\}$ a class of functions in the RKHS $\mathcal{H}$ and by $X_{\mathcal{X}}^{\mathrm{P}} = \left(x_1^s, ..., x_{n_s}^s\right) \sim \mathrm{P}$ and $X_{\mathcal{X}}^{\mathrm{Q}} = \left(x_1^t, ..., x_{n_t}^t\right) \sim \mathrm{Q}$ the source and target dataset, respectively. Let $loss(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \to [0, \Upsilon]$ be a $q-$Lipschitz loss function, i.e., for all $a, b \in \mathcal{Y} \times \mathcal{Y}$, $\left|\left|loss(a) - loss(b)\right|\right| = q\left||a - b|\right|$. Denote by $W_{\mathrm{P}}^*, W_{\mathrm{Q}}^*$ the optimal functions learnt from domain P and domain Q, respectively; then, for any hypothesis $h \in H$, with probability of at least $1 - \delta$, the following generalization bound holds with the Rademacher complexity $\Re_{X_{\mathcal{X}}^{\mathrm{P}}}(H)$ over $X_{\mathcal{X}}^{\mathrm{P}}$:

$$\begin{aligned}&\mathcal{L}_{\mathrm{Q}}\left(h, f_{\mathrm{Q}}\right) - \mathcal{L}_{\mathrm{Q}}\left(h_{\mathrm{Q}}^*, f_{\mathrm{Q}}\right) \leq \mathcal{L}_{\mathrm{P}}\left(h, h_{\mathrm{P}}^*\right) + 2q\Re_{X_{\mathcal{X}}^{\mathrm{P}}}(H)\\ &+3B\sqrt{\frac{\log\frac{2}{\delta}}{2n_t}} + 8r\sqrt{\Xi_{\phi}(\{\mu_{\mathrm{Q}}, \mu_{\mathrm{P}}\})} + \Psi\left(\left\{W_{\mathrm{P}}^*, W_{\mathrm{Q}}^*\right\}\right).\end{aligned} \tag{42}$$

Theorem 5 clearly shows that the projected domain scatter $\Xi_{\phi}(\{\mu_{\mathrm{Q}}, \mu_{\mathrm{P}}\})$ and MAF $\Psi\left(\left\{W_{\mathrm{P}}^*, W_{\mathrm{Q}}^*\right\}\right)$ can control the generalization performance of MACI with its empirical measure, that is, to minimize the PDS (or, alternatively, the distributional scatter discrepancy) and MAF (or model discrimination discrepancy) in our methods can effectively improve the generalization bound in the setting of MDA or domain generalization, which is also supported by the following real-world experiments.

EXPERIMENTS

In this section, to evaluate the effectiveness of MACI for emotion recognition, we compare it with several state-of-the-art methods on two benchmark datasets, i.e., DEAP (Koelstra et al., 2012) and SEED (Zheng and Lu, 2015), which are also widely adopted as benchmark datasets for EEG-based emotion recognition (Mansour et al., 2009). Since existing deep DA models have demonstrated to be very effective, mainly applied to the EEG-based emotion recognition problems (Lotfi and Akbarzadeh, 2014), we divide our experiments into two parts, i.e., comparisons with shallow (traditional) DA methods on those emotion recognition tasks mentioned above and comparisons with the deep (CNN-based) DA methods for EEG-based emotion recognition on several cross-datasets.

DATA PREPARATION

At present, there are some EEG datasets for emotional state research. In this article, we used the following two public datasets: DEAP (Koelstra et al., 2012) and SEED (Zheng and Lu, 2015). As reported in Zhong et al. (2020) and Lan et al. (2018), there is a significant difference between these two databases due to some technical aspects. We also adopted the same feature extraction strategy with that in Lan et al. (2018). More details about these two databases can be found in Lan et al. (2018).

In our experiments, differential entropy (DE) (Zhong et al., 2020) is employed as the feature of emotion recognition. In the literature (Shi et al., 2013; Zheng et al., 2015; Chai et al., 2016; Zheng and Lu, 2016; Chai et al., 2017; Lan et al., 2018;

Zhong et al., 2020) about the DA emotion recognition based on EEG, DE features have been widely used. The details of DE are explained in Lan et al. (2018).

Baseline Setting

We compare our MACI method with the following state-of-the-art (related) baselines for multi-source emotion recognition tasks. Besides this, we also report the emotion recognition results of MACI using several deep features:

- No adaptation baseline FSSL (Yang et al., 2013)
- Multi-kernel adaptation method: FastDAM (Duan et al., 2012c)[3]
- Multi-KT (Tommasi et al., 2014)[4]: according to Tommasi et al. (2014), we here also use the l_2-norm constraint on p in Multi-KT algorithm
- Adaptive SVM: A-SVM (Yang et al., 2007)[5]
- Domain selection machine (DSM) (Duan et al., 2012a)
- Deep DA methods: DAN (Long et al., 2015) and ReverseGrad (Ganin and Lempitsky, 2015).

For the baseline FSSL without adaptation and the multi-source adaptation method A-SVM, we just equally fuse the decision values of all base classifiers with each classifier learned on one source domain[6].

In our MACI, only several vital parameters such as q_1, q_2, λ, α, and β in our model need to be predefined. Considering that parameter determination is a yet unaddressed open issue, we determine these parameters empirically as in our previous works (Tao et al., 2019). The parameters q_1 and q_2 play the same role in optimizing ϑ_a and η_a for preventing the trivial solution of these optimal variables. Since the larger q_1 (or q_2) would lead to the same weights with greater probability, we therefore empirically set $q_1 = q_2 = 2$ in our experiments in terms of the suggestion provided in Hou et al. (2017). Besides this, we discreetly choose the values of λ, α, and β by employing the grid search strategy in a heuristic way. Concretely, these regularization parameters are tuned from $\{10^{-4}, 10^{-3}, ..., 10^3, 10^4\}$. Finally, we search and fine-tune the number of nearest neighbors k in the set {3, 5, 10, 15, 17} for constructing the affinity graph in MACI (also in FSSL). For our algorithm, the maximum iteration number is set as $\tau = 100$.

For those nonlinear learning methods MACI, FastDAM, and Multi-KT, we borrow the Gaussian kernel [i.e., $K_{i,j} = \exp\left(-\sigma \left|\left|x_i - x_j\right|\right|^2\right)$] as the default kernel function, where σ is determined by setting it to be the reciprocal of feature dimension $1/d$. Following the same practice in Duan et al. (2012a), we predefine each source weight $\gamma_i = \frac{\exp(-\delta Dist(X_i^s, X))}{\sum_i \exp(-\delta Dist(X_i^s, X))}$ ($i = 1, ..., S$) in FastDAM, where $\delta = 100$.

Experiment I: Within-Dataset Emotion Recognition

It is worth noting that we may encounter difficulties with different subjects in EEG emotion recognition even if they belong to the same dataset because different subjects may have different EEG feature distributions due to personalized characteristics. Thereby, we may adopt the so-called leave-one-subject-out cross-validation strategy adopted also in Lan et al. (2018) to evaluate the performance of MACI on emotion recognition. Concretely, the left subject from the dataset of interest contributes to the target domain, and other subjects are constructed as the multi-source domains. We evaluate the multi-source adaptation performance of MACI compared with existing state-of-the-arts on SEED and DEAP, respectively.

There are totally 2,340 training data of 13 subjects with 60 data per class and 180 test data of each subject from three classes in DEAP. We extracted 2,775 samples consisting of 925 samples per class per class hour from each one of the 14 subjects in SEED, thus generating 38,850 training data from 14 subjects and 2,775 test samples from one target subject. Note that extant research (Chai et al., 2016, 2017; Zheng and Lu, 2016) have pointed out that it is almost impossible to train the DA methods by exploiting all training data from SEED due to the limitation of computational space. We thereby randomly sample 10% training data from SEED, i.e., 3,885 training data as the final multi-source domain data for all DA methods. We repeat each trial 10 times on SEED, and the final performance is the average of the results of 10 times.

Performance Comparison

We show in **Table 2** the emotion recognition performance of MACI and several baselines on within-DEAP and within-SEED, respectively.

As reported in Lan et al. (2018), the theoretical performance (or chance level) of random guessing is about 33.33%, which could be approached by real chance level when the number of training samples increase to infinity (Lan et al., 2018). As shown in **Table 1**, the baseline FSSL contributes 40.17% mean recognition accuracy on DEAP, which is very near to the random value. When there are finite samples, we obtain the empirical chance level by repeating the trials of the samples in question equipped with randomized class labels (Lan et al., 2018). The finally obtained chance levels with bound of 95% confidence interval are also recorded in **Table 2**. We can see from **Table 2** that the accuracy of FSSL significantly exceeds the upper bound of the real chance level at 5% significance level. However, the relatively lower performance of FSSL still indicates that emotion recognition with DA technique is imperative when there exists substantial divergence between the feature distributions of different subjects.

Almost all DA methods yield better recognition performance than FSSL for DEAP. Our MACI achieves the best performance (about 23.14% gains in performance over FSSL), closely followed by DSM. Note that though we acquired the relatively significant improvement measured by t-test with p-value > 0.05, the total recognition accuracy is still inferior. On SEED, FSSL achieves

[3]The code is available from http://vc.sce.ntu.edu.sg/transfer_learning_domain_adaptation_data/DAM-TNNLS2012.html.

[4]We use the code available from http://homes.esat.kuleuven.be/~ttommasi/source_code_CVPR10.html.

[5]The MATLAB code is available online http://www.robots.ox.ac.uk/~{}vgg/software/tabularasa/.

[6]For each source domain, we train one SVM by using the corresponding labeled samples. Then, for each test instance x, the decision values from p SVM classifiers are converted into the probability values by using the sigmoid function [i.e., $g(t) = 1/(1 + \exp(-t))$]. Finally, we average the p probability values as the final prediction of the test instance x.

TABLE 2 | Emotion recognition performance (mean % and SD %) of MACI and several baselines on within-datasets.

Method	DEAP		SEED							
			Session I		Session II		Session III		Average	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
FSSL	40.17	4.36	57.96	6.85	48.79	5.47	57.45	9.09	53.78	6.96
Multi-KT	55.83	5.59	73.56	4.37	68.89	3.43	72.57	7.38	70.68	5.09
A-SVM	49.49	7.92	65.82	7.86	64.00	7.09	69.08	10.77	65.25	8.53
FastDAM	57.37	5.50	72.31	6.86	69.45	7.18	75.64	7.37	71.52	7.04
DSM	60.22	6.50	72.76	6.86	70.10	5.18	76.35	7.37	72.27	6.47
MACI	63.31	4.50	73.42	6.86	70.81	6.18	77.43	7.37	73.23	6.66
Upp Bnd of Chn Lvl	38.85		34.58		34.65		34.60		34.61	

53.78% average recognition accuracy over three sessions, which obviously surpasses the upper bound of the chance level. Several multi-source adaptation methods, i.e., Multi-KT, FastDAM, and DSM, undoubtedly obtain more performance gains than FSSL on SEED. The proposed method MACI still effectively boost the mean recognition accuracy up to 73.23% under t-test with p-value > 0.05, which still demonstrates the best performance on SEED. It is worthy to note that all methods including FSSL work more effectively on SEED than on DEAP. This interesting observation is partially consistent with that in Lan et al. (2018). A possible explanation may be that the so-called negative transfer prevented the effective application of DA techniques in DEAP since larger discrepancies among different subjects may exist in DEAP than in SEED (Mansour et al., 2009; Lan et al., 2018).

Number of Source Samples

Figure 1 presents the effect of varying the number of source samples. The source dataset size varies from 100 to 2,300 on DEAP and 100 to 3,800 on SEED, respectively. It can be seen from the curves in **Figure 1** that all methods manifest the same trend of upgrade in the figure. This shows that larger source data is beneficial to improve the learning performance. It is worthy to note that the performance of MACI can be smoothly and steadily improved with the increase of the source samples, while other DA methods may only achieve satisfactory performance when the source samples are relatively large, i.e., larger than 500. In addition, A-SVM obtains the least performance on two datasets due to the so-called negative transfer issue in DA.

On DEAP, MACI, and DSM maintain a better accuracy than other methods with less than 500 source domain samples. DSM outperforms Multi-KT and FastDAM when the number of target training samples is relatively large due to properly choosing the weights to assign to each subject. Our method MACI obtains even more gains over DSM when the number of source samples is increasing asymptotically. The reason may be that, except for the use of correlation information among subjects, MACI can effectively select the most related sources with the optimally weighted multi-source adaptation regularization. Besides that, partial experimental results also show that FastDAM could occasion the "negative transfer" issue with the MMD-based weights assigned to all sources, which would deteriorate its performance. On SEED, the accuracy flattens at above 3,000 source samples. When the number of source samples increases to 3,500, MACI, FastDAM, and Multi-KT asymptotically approach a similar performance. From this point onwards, MACI, FastDAM, and Multi-KT perform similarly if we have sufficient source data.

Multi-Kernel Learning

We further evaluate the effectiveness of our method with different kernel functions (called MKMACI for short) for each source domain. Given the empirical kernel mapping set $\{\phi_a\}_{a=1}^{\upsilon}$, each mapping X_a into a different kernel space, we can integrate them orthogonally to the final space by concatenation, i.e., $\widetilde{\phi}(x_i) = [\phi_1(x_i)^T, \phi_2(x_i)^T, ..., \phi_\upsilon(x_i)^T]^T \in R^{\upsilon n_a}$, for $x_i \in X_a$. The final kernel matrix in this new space is defined as $K_{new} = [\mathring{\mathrm{K}}_1; \mathring{\mathrm{K}}_2; ...; \mathring{\mathrm{K}}_\upsilon]$, where $\mathring{\mathrm{K}}_i$ is the kernel matrix in the $i-$th feature space. Therefore, besides the above-mentioned Gaussian kernel, we additionally employ another three types of kernels in MKMACI: Laplacian kernel $K_{ij} = \exp\left(-\sqrt{\sigma}\left||x_i - x_j\right|\right)$, inverse square distance kernel $K_{ij} = 1\Big/\left(1 + \sigma\left||x_i - x_j\right|\right|^2\right)$, and inverse distance kernel $K_{ij} = 1/(1 + \sqrt{\sigma}\left||x_i - x_j\right|\right)$. It can be clearly seen in **Figure 2** that MKMACI is obviously better than MACI in terms of mean accuracies in all cases, which justifies that the multi-kernel trick can improve the quality of DA emotion recognition on within-datasets.

Experiment II: Cross-Dataset Emotion Recognition

Note that cross-dataset emotion recognition is more challenging in terms of the differences in acquisition and participant characteristics and behaviors. In the preceding experiments, we demonstrate the performance comparison of our method with other DA methods with the within-dataset (i.e., cross-subject) setting. We will, in this part, further evaluate the consistent effectiveness of MACI when performed on cross-dataset adaptation. In this scenario of experiment, we constructed multiple different schemes by sampling the training dataset and test dataset, respectively, with different EEG instruments and emotional stimuli. We therefore set up six trial settings, i.e., *DEAP* → *SEED I*, *DEAP* → *SEED II*, *DEAP* → *SEED III*, *SEED I* → *DEAP*, *SEED II* → *DEAP*, and *SEED III* → *DEAP*, to justify the effectiveness of MAC on cross-dataset emotion

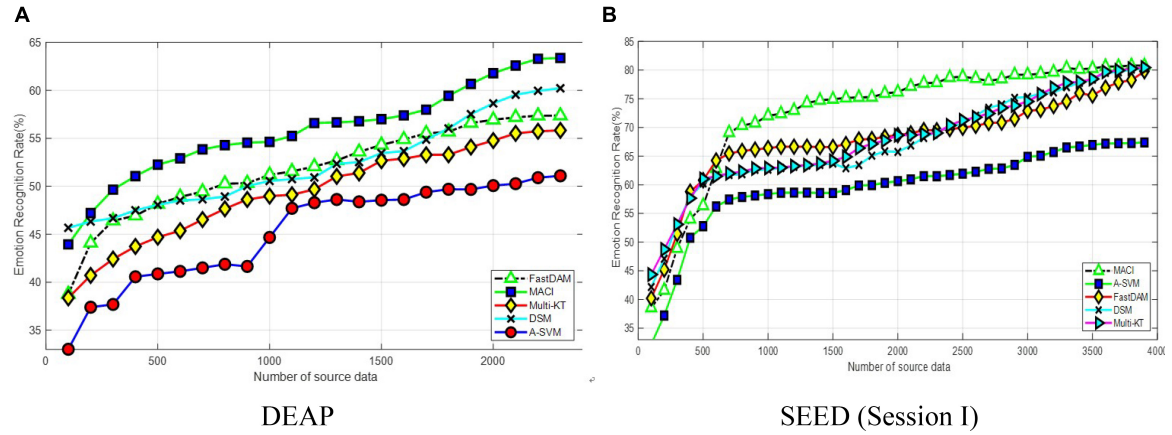


FIGURE 1 | Classification accuracy with varying numbers of source samples on **(A)** DEAP and **(B)** SEED (session I).

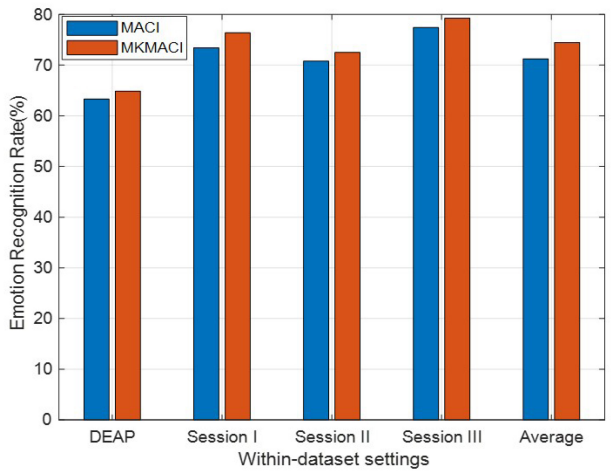


FIGURE 2 | Domain adaptation emotion recognition on within-dataset with multi-kernel learning.

recognition. In the context, we denote A→B by adaptation from dataset A to dataset B. For simplicity of expression, we respectively, coined SEED I, SEED II, and SEED III as the dataset of session I, session II, and session III in the database SEED.

In universe DA, a commonly used hypothesis is that the feature space of both source and target domains is the same. Consequently, only 32 channels between SEED and DEAP are employed to formulate a 160 dimensional feature space for both training and test datasets. In the first three experimental settings, there are 180 × 14 = 2,520 source samples from DEAP and 2,775 target samples from three different sessions in SEED. We evaluate the recognition accuracy for each subject in each session and report the final experimental results based on the mean over 15 subjects from SEED. In the other experimental settings, a total of 2,775 × 15 = 41,625 source samples from SEED are regarded as training datasets, and 180 samples contributed from DEAP are test dataset. We then evaluate the recognition accuracy of individual subjects in DEAP, and the results are recorded with the average over 14 subjects. We randomly sample 10% of the source data (4,162 samples) as the actual training data due to the limitation of memory (Shi et al., 2013; Zheng et al., 2015; Chai et al., 2016, 2017; Zheng and Lu, 2016; Lan et al., 2018; Zhong et al., 2020). Under each setting, we conduct the trial repeatedly 10 times and record the average performance over these 10 times.

Performance Comparison

We record the mean experimental results on six cross-dataset settings in **Table 3**, from which we can observe that the performance of the baseline FSSL is inferior to the upper bound of chance level with 95% confidence interval, that is, the baseline performance is almost close to the random guess with 5% significance level. This indicates that there exist larger distribution divergences between two datasets as well as the variance among different subjects than that in within-dataset. The importance of DA would be indispensable in this scenario. This is justified by the observation in **Table 3** that all DA methods outperform the baseline FSSL since DA could potentially reduce the technical discrepancies in cross-dataset applications. In most cases, MACI is found to be the best-performing method in the cross-dataset DA settings. In some scenario, Multi-KT and FastDAM occasionally obtain the best performance. A noticeable phenomenon can be observed in **Table 3**, such that the mean recognition accuracies of all methods are correspondingly worse than that in **Table 2** obtained on within-dataset due to the larger distribution discrepancy between different datasets.

Multi-Source Adaptation

In practical DA applications, one may expect that the number of prior sources grow in time, which would incur the so-called scalability issue. In this problem, it is necessary to explore the reliability of each prior source for the specific task (Tao et al., 2019). To this end, we additionally conduct multi-source adaptation trials on several cross-dataset settings. The average results of MACI, DSM, Multi-KT, FastDAM, and A-SVM with the average prior model are reported in **Figure 3**.

TABLE 3 | The recognition accuracy (mean%) with cross-dataset settings.

Method	DEAP→SEED I	DEAP→SEED II	DEAP→SEED III	SEED I→DEAP	SEED II→DEAP	SEED III→DEAP
FSSL	32.42	33.71	34.47	33.57	32.99	32.51
A-SVM	55.86	58.48	60.84	39.68	40.08	39.53
FastDAM	65.72	62.68	66.21	48.40	49.90	47.46
DSM	68.47	64.68	64.33	50.22	51.44	50.46
Multi-KT	67.74	65.51	64.65	48.73	52.16	51.27
MACI	69.36	67.60	65.43	54.37	51.88	51.76
Upp Bnd of Chn Lvl.	34.68	34.72	34.74	38.35	38.38	38.44

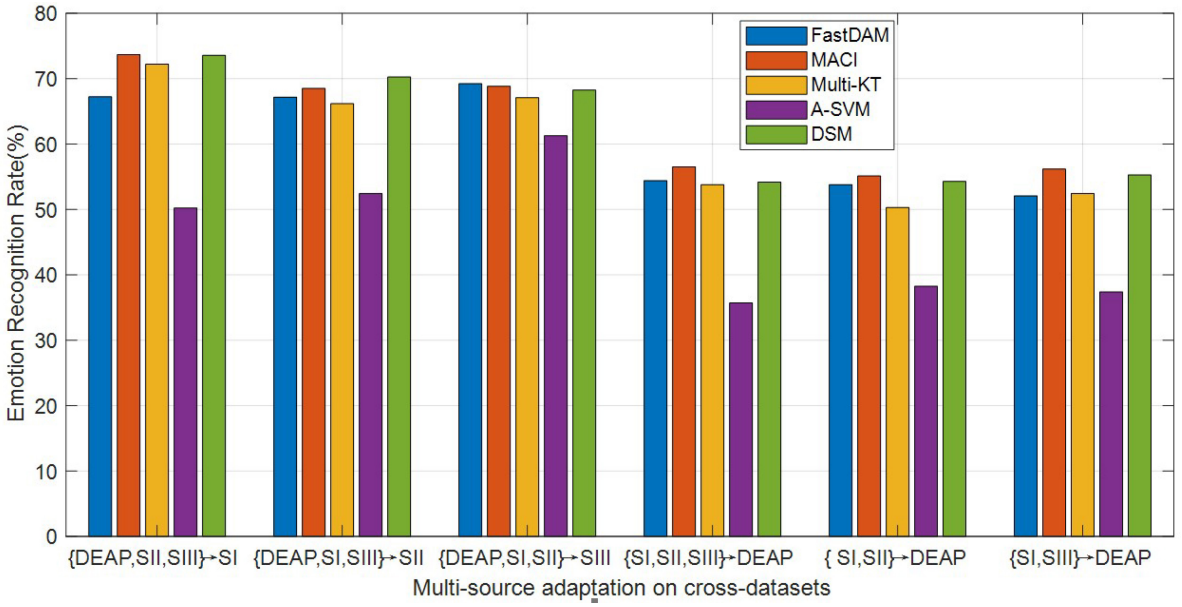


FIGURE 3 | Multi-source adaptation emotion recognition accuracy (SI, session I; SII, session II; SIII, session III).

TABLE 4 | Cross-dataset emotion recognition rates with different strategies of parameter settings.

Method	{DEAP, SII, SIII}→SI	{DEAP, SI, SIII}→SII	{DEAP, SI, SII}→SIII	{SI, SII, SIII}→DEAP	{SI, SII}→DEAP	{SI, SIII}→DEAP
MACI_NF	73.32	67.24	69.78	54.62	55.04	54.39
MACI_NS	69.88	65.31	66.07	52.92	53.44	52.81
MACI	73.69	68.52	68.85	56.52	55.12	56.17

SI, session I; SII, session II; SIII, session III.

It can be seen from the curves in **Figure 3** that A-SVM is still worse than the other DA methods in most cases in that it is difficult for A-SVM to minimize the between-domain distribution distance when the distribution varies greatly between domains. The accuracies of A-SVM tend to be downgraded when the number of sources is increasing in some cases, suggesting that negative transfer may have happened in A-SVM. MACI obtains a relatively much better performance in most cases, which demonstrates that our algorithm can improve the emotion recognition performance on cross-dataset. All methods except A-SVM manifest the same trend of upgrade with the increase of sources, and the accuracy improvements are significant with respect to that of within-dataset settings. This shows that utilizing the limited sources is beneficial to improve the learning performance. In addition, MACI and DSM usually outperform other DA methods due to properly choosing the weights to assign to each source. Our method MACI obtains even more gains over DSM, which may be attributed to the utilization of correlation information among sources in that MACI can effectively select the most related source domains with the optimally weighted multi-source adaptation regularization.

Adaptation With Deep Features

In the past decade, deep learning attracts more and more attention due to its powerful representation ability and dramatic improvement over the traditional shallow methods. We therefore additionally compare our MACI method with the recently proposed deep transfer learning models DAN and ReverseGrad

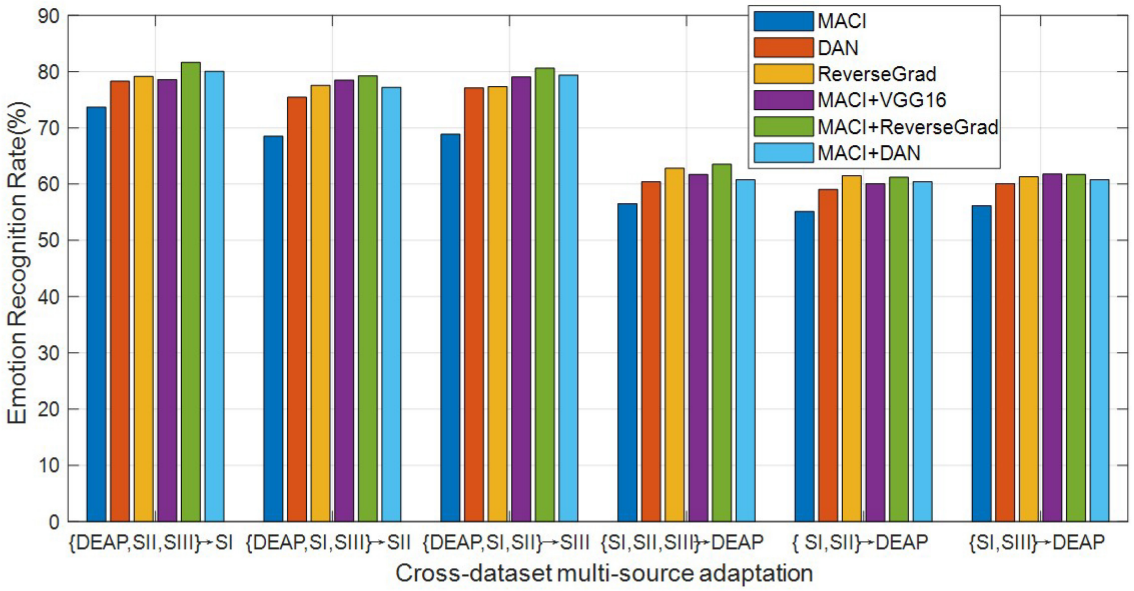


FIGURE 4 | Emotion recognition accuracies (%) of different methods using deeply extracted features (SI, session I; SII: session II; SIII, session III).

for cross-dataset emotion recognition using deeply extracted features in multi-source adaptation settings.

In our MACI, we can tackle the problem of deep DA with two steps: firstly, a higher-level feature extraction is learnt in an unsupervised fashion from all available domains using the popular deep architectures [e.g., VGG16 (Simonyan and Zisserman, 2014) or DAN]; secondly, our MACI is trained on the transformed data of all domains and then used to test the target domain. For fair comparison, however, we follow the experimental setup in Zhou et al. (2018) and Zhu et al. (2017). Specifically, we first fine-tune pretrained deep models (e.g., VGG16, DAN, and ReverseGrad) by using the labeled samples in the source domain and then use these fine-tuned CNN models to extract the features from EEG in both source and target domains. Finally, we perform emotion recognition using MACI on these deeply extracted features. In the context of our experiments, we denote our methods with different deep models as MACI + VGG16, MACI + DAN, and MACI + ReverseGrad, respectively. As for DAN and ReverseGrad, we use their released source codes and fine-tune the pre-trained deep models by using the suggested parameters in Long et al. (2015) and Ganin and Lempitsky (2015), respectively.

All experimental results are reported in **Figure 4**. As can be seen from this plot, the deep transfer learning methods are originally proposed to learn domain-invariant features, while our proposed method aims to improve the cross-domain generalization ability, namely, their methods focus on feature learning, while our work focuses on classification, so our proposed method can be used to further improve the recognition accuracies by co-learning the source classifiers with the features extracted by deep models, i.e., VGG16, DAN, and ReveseGrad. This indicates that the classification-level constraint can preserve all source discriminative structures for the guidance of target data classification, which demonstrates the effectiveness of MACI framework. From the plot bars of **Figure 4**, it can be observed that MACI + DAN consistently outperforms DAN, while MACI + ReverseGrad is consistently better than ReveseGrad, which demonstrates that our MACI method is complementary to the two deep transfer learning methods DAN and ReveseGrad by exploiting the correlation statistics to further enhance the generalization ability across domains.

Parameter Impact on MACI

There are mainly three model parameters to be tuned in our method, i.e., λ, β, and α. Note that larger α would make the predicted label matrix better meet the expected needs, thus with better results being achieved. Consequently, we empirically set $\alpha = 10^3$ in the following experiments. We firstly explore to set the extreme values of different parameters for validating the importance of each component in our framework. Specifically, we denote MACI without the feature selection (i.e., $\beta = 0$) by MACI_NF and MACI with $\lambda = \eta_a = 0$ by MACI_NS, which ignores correlation information among multiple sources. These settings are evaluated on cross-dataset settings for multi-source adaptation tasks. From **Table 4**, we can observe that MACI can be significantly improved from MACI_NS by exploiting the correlation information among multiple sources. Besides this, the performance of MACI_NF is slightly weaker than MACI, that is, MACI would degrade when the feature selection function is omitted. A possible reason may be that the features of EEG represented by DE introduced some noise/outlier data. In this case, the feature selection in MACI possesses indispensable importance for robust DA learning. In sum, the utilization of correlation knowledge among sources and features could make

MACI further boost its performance in cross-dataset emotion recognition applications. It is this argument that constitutes the basic principle of our MACI framework.

CONCLUSION

In this work, we explore to cope with the cross-dataset emotion recognition where existing BCI methods cannot work well. To this end, we proposed an effective multi-source co-adaptation framework (MACI) for EEG-based emotion recognition mainly by leveraging correlation knowledge among sources and features in the objective function, which dampens unimportant evidence (within features and between sources) and amplifies useful knowledge. In MACI, multiple domain-invariant classification functions corresponding to different sources are co-learned by bridging both statistical and semantic distribution discrepancy between source and target domains, thus making MACI utilize the correlated knowledge among multiple sources by exploiting the developed correlation metric function. A large number of experimental results conducted on two publicly available EEG datasets show that MACI are much better than several representative baseline methods and provide the state-of-the-art performance on within/cross-dataset emotion recognition in most cases. This demonstrates the effectiveness of MACI in addressing feature distribution discrepancy between individual subjects as well as different datasets due to technical discrepancies.

To boost the efficiency of our method, however, a more efficient iterative algorithm would be developed or further elaborated in our future works. Besides this, the pseudo-labels strategy (i.e., iteratively updating target label matrix in the training stage) for bridging semantic distribution discrepancy between different domains would be unreliable or even misleading in training. This therefore arouses another challenge, i.e., how to effectively infer and incorporate target labels in unsupervised DA, which would be an urgent and valuable work in our future research.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html.

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.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported partly by High level talents introduction research project of Ningbo Polytechnic (under grant no. RC201902), partly by Zhejiang Provincial Natural Science Foundation of China (under grant no. LY19F020012), and partly by Foundation of Zhejiang Educational Committee (under grant no. Y201941140).

REFERENCES

Bruzzone, L., and Marconcini, M. (2010). Domain adaptation problems: a DASVM classification technique and a circular validation strategy. *IEEE Trans. PAMI* 32, 770–787. doi: 10.1109/tpami.2009.57

Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., et al. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* 17:1014. doi: 10.3390/s17051014

Chai, X., Wang, Q., Zhao, Y., Liu, X., Bai, O., and Li, Y. (2016). Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. *Comput. Biol. Med.* 79, 205–214. doi: 10.1016/j.compbiomed.2016.10.019

Chen, B., Lam, W., Tsang, I. W., and Wong, T.-L. (2013). Discovering low-rank shared concept space for adapting text mining models. *IEEE Transact. Patt. Anal. Mach. Intell.* 35, 1284–1297. doi: 10.1109/tpami.2012.243

Chu, W. S., Del, T. F., and Cohn, J. F. (2017). Selective transfer machine for personalized facial action unit detection. *IEEE Transact. Patt. Anal. Mach. Intell.* 39, 529–545.

Ding, Z., Sheng, L., Ming, S., and Fu, Y. (2018). "Graph adaptive knowledge transfer for unsupervised domain adaptation," in *Proceedings of the 15th European Conference (ECCV2018), Munich, Germany, September 8-14, 2018*, (Cham: Springer).

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194.

Duan, L., Dong, X., and Shih-Fu, C. (2012a). "Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach," in *Proceedings of the CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, (New York, NY: IEEE), 1338–1345.

Duan, L., Tsang, I. W., and Xu, D. (2012b). Domain transfer multiple kernel learning. *IEEE Transact. Patt. Anal. Mach. Intell.* 34, 465–479. doi: 10.1109/tpami.2011.114

Duan, L., Xu, D., and Tsang, I. W. (2012c). Domain adaptation from multiple sources: a domain-dependent regularization approach. *IEEE Transact. Neur. Netw. Learn. Syst.* 23, 504–518. doi: 10.1109/tnnls.2011.2178556

Ganin, Y., and Lempitsky, V. (2015). "Unsupervised domain adaptation by back-propagation," in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, 1180–1189.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Lavi-olette, F., et al. (2016). Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* 17, 1–35. doi: 10.1109/tnnls.2020.3025954

Ghifary, M., Balduzzi, D., Kleijn, W. B., and Zhang, M. (2017). Scatter component analysis: a unified framework for domain adaptation and domain generalization. *IEEE Transact. Patt. Anal. Mach. Intell.* 99, 1–1.

Gretton, A., Fukumizu, K., Harchaoui, Z., and Sriperumbudur, B. K. (2009). "A fast, consistent kernel two-sample test," in *Proceedings of the Conference on Neural Information Processing Systems 22*, (Vancouver, BC: MIT Press), 673–681.

Hou, C., Feiping, N., Hong, T., and Yi, D. (2017). Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Trans. Knowl. Data Eng.* 29, 1998–2011. doi: 10.1109/tkde.2017.2681670

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Magaz.* 11, 20–31.

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transact. Aff. Comput.* 5, 327–339. doi: 10.1109/taffc.2014.2339834

Judy, H., Tzeng, E., Darrell, T., and Saenko, K. (2017). Simultaneous deep transfer across domains and tasks. *Dom. Adaptat. Comput. Vis. Appl.* 17, 173–187. doi: 10.1007/978-3-319-58347-1_9

Kim, M.-K., Kim, M., Oh, E., and Kim, S.-P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Comput. Mathematic. Methods Med.* 2013, 1–3. doi: 10.1155/2013/573734

Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis using physiological signals. *IEEE Transact. Affect. Comput.* 3, 18–31. doi: 10.1109/t-affc.2011.15

Lan, Z., Sourina, O., Wang, L., Scherer, R., and Müller-Putz, G. R. (2018). Domain adaptation techniques for eeg-based emotion recognition: a comparative study on two public datasets. *IEEE Transact. Cogn. Dev. Syst.* 11, 85–94. doi: 10.1109/tcds.2018.2826840

Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., and Hu, B. (2018a). Exploring EEG features in cross-subject emotion recognition. *Front. Neurosci.* 12:162.

Li, Y., Zheng, W., Cui, Z., Zhang, T., and Zong, Y. (2018b). "A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, Stockholm, 1561–1567.

Li, Y., Zheng, W., Cui, Z., Zong, Y., and Ge, S. (2018c). EEG emotion recognition based on graph regularized sparse linear regression. *Neur. Proces. Lett.* 8, 1–17. doi: 10.1109/taffc.2020.2994159

Li, Z., Liu, J., Tang, J., and Lu, H. (2015). Robust structured subspace learning for data representation. *IEEE Transact. Patt. Anal. Mach. Intell.* 37, 2085–2098. doi: 10.1109/tpami.2015.2400461

Long, M., Cao, Y., Wang, J., and Jordan, M. (2015). "Learning transferable features with deep adaptation networks," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Lille, 97–105.

Long, M., Wang, J., Ding, G., and Pan, S. J. (2014). Adaptation regularization: a general framework for transfer learning. *IEEE Transact. Knowl. Data Eng.* 26, 1076–1089. doi: 10.1109/tkde.2013.111

Lotfi, E., and Akbarzadeh, M.-R. (2014). Practical emotional neural networks. *Neur. Netw.* 59, 61–72. doi: 10.1016/j.neunet.2014.06.012

Mansour, Y., Mohri, M., and Rostamizadeh, A. (2009). "Domain adaptation with multiple sources," in *Proceedings of the Conference on Neural Information Processing Systems*, (Vancouver, BC: MIT Press), 1041–1048.

Mühl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain Comput. Interfac.* 1, 66–84. doi: 10.1080/2326263x.2014.912881

Musha, T., Terasaki, Y., Haque, H. A., and Ivamitsky, G. A. (1997). Feature extraction from EEGs associated with emotions. *Artif. Life Robot.* 1, 15–19. doi: 10.1007/bf02471106

Nie, F., Huang, H., Cai, X., and Ding, C. (2010a). "Efficient and robust feature selection via joint -norms minimization," in *Proceedings of the International Conference on Neural Information Processing Systems*, (Vancouver, BC: Curran Associates Inc), 1813–1821.

Nie, F., Xu, D., Tsang, I. V.-H., and Zhang, C. (2010b). Flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction. *IEEE Transact. Image Proces.* 19, 1921–1932. doi: 10.1109/tip.2010.2044958

Pan, S. J., Tsang, I. W., Kwok, J. T., and Yang, Q. (2011). Domain adaptation via transfer component analysis. *IEEE Transact. Neural Netw.* 22, 199–210. doi: 10.1109/tnn.2010.2091281

Pandey, P., and Seeja, K. (2019). "Emotional state recognition with EEG signals using subject independent approach," in *Data Science and Big Data Analytics. Lecture Notes on Data Engineering and Communications Technologies*, Vol. 16, eds D. Mishra, X. S. Yang, and A. Unal (Singapore: Springer), 117–124. doi: 10.1007/978-981-10-7641-1_10

Rosenstein, M. T., Marx, Z., and Kaelbling, L. P. (2005). "To Transfer or not to transfer," in *Proceedings of the Conference on Neural Information Processing Systems*, (Cambridge, MA: MIT Press).

Shi, L. C., Jiao, Y. Y., and Lu, B. L. (2013). "Differential entropy feature for EEG-based vigilance estimation," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, 6627–6630.

Simonyan, K., and Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference Learning Representations (ICLR)*, Banff, 1–14.

Song, T., Zheng, W., Song, P., and Cui, Z. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transact. Affect. Comput.* 18, 1–1.

Tao, J. W., Song, D., Wen, S., and Hu, W. (2017). Robust multi-source adaptation visual classification using supervised low-rank representation. *Patt. Recogn.* 61, 47–65. doi: 10.1016/j.patcog.2016.07.006

Tao, J., Chung, F. L., and Wang, S. (2012). On minimum distribution discrepancy support vector machine for domain adaptation. *Patt. Recogn.* 45, 3962–3984. doi: 10.1016/j.patcog.2012.04.014

Tao, J., Di, Z., Fangyu, L., and Bin, Z. (2019). Latent multi-feature co-regression for visual recognition by discriminatively leveraging multi-source models. *Patt. Recogn.* 87, 296–316. doi: 10.1016/j.patcog.2018.10.023

Tao, J., Wen, S., and Hu, W. (2015). L1-norm locally linear representation regularization multi-source adaptation learning. *Neural Netw.* 69, 80–98. doi: 10.1016/j.neunet.2015.01.009

Tao, J., Wen, S., and Hu, W. (2016). Multi-source adaptation learning with global and local regularization by exploiting joint kernel sparse representation. *Knowl. Based Syst.* 98, 76–94. doi: 10.1016/j.knosys.2016.01.021

Tommasi, T., Orabona, F., and Caputo, B. (2014). Learning categories from few examples with multi model knowledge transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 928–941. doi: 10.1109/tpami.2013.197

Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017). "Adversarial discriminative domain adaptation," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2962–2971.

Wang, C., and Mahadevan, S. (2011). "Heterogeneous domain adaptation using manifold alignment," in *Proceedings of the International Joint Conference on Artificial Intelligence*, Barcelona, 1541–1546.

Yan, S., Xu, D., Zhang, B., Yang, Q., and Lin, S. (2006). Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Transact. Patt. Anal. Mach. Intell.* 29:40. doi: 10.1109/tpami.2007.250598

Yang, J., Yan, R., and Hauptmann, A. G. (2007). "Cross-domain video concept detection using adaptive svms," in *Proceedings of the ACM International Conference on Multimedia*, (New York, NY: ACM), 188–197.

Yang, Y., Ma, Z., Hauptmann, A. G., and Sebe, N. (2013). Feature selection for multimedia analysis by sharing information among multiple tasks. *IEEE Transact. Multim.* 15, 661–669. doi: 10.1109/tmm.2012.2237023

Zhang, K., Gong, M., and Schölkopf, B. (2015). "Multi-source domain adaptation: a causal view," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, Austin, TX, 3150–3157.

Zhang, Y., Chung, F. L., and Wang, S. (2019a). Takagi-sugeno-kang fuzzy systems with dynamic rule weights. *J. Intell. Fuzzy Syst.* 37, 8535–8550. doi: 10.3233/jifs-182561

Zhang, Y., Chung, F., and Wang, S. (2020a). Clustering by transmission learning from data density to label manifold with statistical diffusion. *Knowl Based Syst.* 193:105330. doi: 10.1016/j.knosys.2019.105330

Zhang, Y., Dong, J., Zhu, J., and Wu, C. (2019b). Common and special knowledge-driven TSK fuzzy system and its modeling and application for epileptic EEG signals recognition. *IEEE Access* 7, 127600–127614. doi: 10.1109/access.2019.2937657

Zhang, Y., Li, J., Zhou, X., Zhang, M., Ren, J., and Yang, J. (2019c). A view-reduction based multi-view TSK fuzzy system and its application for textile color classification. *J. Amb. Intell. Human. Comput.* 19, 1–11.

Zhang, Y., Tian, F., Wu, H., Xingyun, G., Danmin, Q., Jiancheng, D., et al. (2017). Brain MRI tissue classification based fuzzy clustering with competitive learning. *J. Med. Imag. Health Inform.* 7, 1654–1659. doi: 10.1166/jmihi.2017.2181

Zhang, Y., Wang, L., Wu, H., Geng, X., Yao, D., and Dong, J. (2016). A clustering method based on fast exemplar finding and its application on brain magnetic resonance images segmentation. *J. Med. Imag. Health Inform.* 6, 1337–1344. doi: 10.1166/jmihi.2016.1923

Zhang, Y., Wang, S., Xia, K., Jiang, Y., and Qian, P. (2020b). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inform. Fus.* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002

Zheng, W. (2017). Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. *IEEE Transact. Cogn. Dev. Syst.* 9, 281–290. doi: 10.1109/tcds.2016.2587290

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transact. Autonom. Mental Dev.* 7, 162–175. doi: 10.1109/tamd.2015.2431497

Zheng, W. L., and Lu, B. L. (2016). "Personalizing EEG-based affective models with transfer learning," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, Palo Alto, CA, 2732–2738.

Zheng, W. L., Zhang, Y. Q., Zhu, J. Y., and Lu, B. L. (2015). "Transfer components between subjects for EEG-based emotion recognition," in *Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Xi'an, 917–922.

Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Transact. Affect. Comput.* 99, 1–1.

Zhou, X., Jin, K., Shang, Y., and Guo, G. (2018). Visually interpretable representation learning for depression recognition from facial Im-ages. *IEEE Transact. Affect. Comput.* 11, 542–552. doi: 10.1109/TAFFC.2018.2828819

Zhu, Y., Shang, Y., Shao, Z., and Guo, G. (2017). Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Transact. Affect. Comput.* 9, 578–584. doi: 10.1109/TAFFC.2017.2650899 doi: 10.1109/taffc.2017.2650899

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fnins.2021.687496

A Transfer Model Based on Supervised Multi-Layer Dictionary Learning for Brain Tumor MRI Image Recognition

Yi Gu and Kang Li*

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China

Artificial intelligence (AI) is an effective technology for automatic brain tumor MRI image recognition. The training of an AI model requires a large number of labeled data, but medical data needs to be labeled by professional clinicians, which makes data collection complex and expensive. Moreover, a traditional AI model requires that the training data and test data must follow the independent and identically distributed. To solve this problem, we propose a transfer model based on supervised multi-layer dictionary learning (TSMDL) for brain tumor MRI image recognition in this paper. With the help of the knowledge learned from related domains, the goal of this model is to solve the task of transfer learning where the target domain has only a small number of labeled samples. Based on the framework of multi-layer dictionary learning, the proposed model learns the common shared dictionary of source and target domains in each layer to explore the intrinsic connections and shared information between different domains. At the same time, by making full use of the label information of samples, the Laplacian regularization term is introduced to make the dictionary coding of similar samples as close as possible and the dictionary coding of different class samples as different as possible. The recognition experiments on brain MRI image datasets REMBRANDT and Figshare show that the model performs better than competitive state of-the-art methods.

Keywords: brain tumor MRI image, supervised learning, transfer learning, Laplacian regularization, multi-layer dictionary learning

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Yanhui Zhang,
Hebei University of Chinese Medicine, China
Hongru Zhao,
Soochow University, China

****Correspondence:***
Yi Gu
8202101437@jiangnan.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *29 March 2021*
Accepted: *19 April 2021*
Published: *28 May 2021*

Citation:
Gu Y and Li K (2021) A Transfer Model Based on Supervised Multi-Layer Dictionary Learning for Brain Tumor MRI Image Recognition. Front. Neurosci. 15:687496. doi: 10.3389/fnins.2021.687496

INTRODUCTION

Brain tumor is a common neurological disease. As a high incidence disease, its incidence rate has reached 1.34 per 100,000 in China, and over 200,000 patients diagnosed with primary or metastatic brain tumors in the United States every year. Among the incidence of systemic tumors, brain tumors are second only to those of the stomach, uterus, breast, and esophagus, accounting for approximately 2% of systemic tumors and the proportion of deaths has exceeded 2% (Sun et al., 2019; Sung et al., 2021). According to surveys, the incidence rate of brain tumors is the highest among children, and the highest incidence is 20–50-year-old young adults. Among childhood malignancies, brain tumors are the second most common, after leukemia. Brain tumors not only cause physical and mental suffering to patients, but also place a heavy financial burden on their families. As a standard technique for non-invasive brain tumor diagnosis, magnetic resonance

imaging (MRI) is an essential component of medical diagnosis and treatment. It uses magnetic resonance phenomena to obtain electromagnetic signals from the brain, so as to reconstruct brain information and provide a validated anatomical image of the brain. MRI can increase the diagnostic ability of medical diagnosticians. The wide application of MRI mainly benefits from the following characteristics (Amin et al., 2017; Bahadure et al., 2017): (1) no bony artifacts, good soft tissue resolution and clear visualization of soft tissue structures; (2) ability to image multiple aspects and multiple parameters, facilitating the acquisition of diagnostic information as a means of determining the various characteristics of the lesion; (3) no radiological damage and no ionizing radiation damage; (4) different profiles can be selected by adjusting the magnetic field, resulting in a three-dimensional image with different angles, which facilitates the identification of the lesion site; (5) has a flow-space effect and does not require an external contrast agent, allowing direct visualization of the vascular structure and facilitating the observation of the relationship between the vessel and the lesion. However, it is time consuming for radiologists to interpret the large number of MRI images and detect early brain tumors. These medical images need to be analyzed by doctors one by one, and the condition should be determined according to their experience.

Artificial intelligence (AI) technology, especially in particular medical image processing, is an effective way to address this challenge (Zeng et al., 2018; Sajjad et al., 2019; Mittal et al., 2019; Ge et al., 2020; Hua et al., 2021). In the process of brain disease diagnosis, firstly, the image features are extracted, and then the extracted image features are classified to complete the image classification and recognition. For example, Ismael and Abdel-Qader (2018) used Gabor filter and discrete wavelet transform to extract statistical features for brain tumor classification. Then this method used the tumor segmented as input and multi-layer perceptron (MLP) as the classifier. Liu et al. (2012) proposed a multi-level classification method for meningiomas. According to the type and growth rate of tumors, meningiomas are divided into three levels. In the classification step, the authors used a multiple logistic regression model. Mallick et al. (2019) proposed a brain MRI image classification method based on deep neural networks. Using encoding and decoding techniques, this method mainly used an automatic autoencoder to extract and classify brain images. To assist radiologists in MRI classification, Sachdeva et al. (2016) proposed a semi-automated classification method with multiple stages. To detect tumor regions, the first stage was the outline system detection of the content-based tumor regions, which can be manually indicated by the physician, called segmented regions of interest (SROI). Then, 71 texture and intensity features were extracted from the SROI regions, and the features were optimized by genetic algorithm. In the classification stage, support vector machine (SVM) and artificial neural network were used. Nikam and Shinde (2013) proposed a brain MRI image classification method based on distance learning. Firstly, the images were preprocessed, and many techniques such as gray transformation, median filtering, and high pass filtering were used to remove the noise of MRI brain image. The threshold segmentation method was used to segment the MRI brain image. Then the features are extracted by correlation, entropy, contrast, homogeneity, and energy. Finally, a Euclidean distance classifier was used for classification. Ghassemi et al. (2020) proposed a CNN model for multi-class brain tumor classification. Firstly, the method was pre-trained as a discriminator in generative adversarial network to extract image features. Second, the softmax classifier was used to distinguish the three kinds of tumors. This model consists of six layers, which can be used together with various data augmentation techniques. Kiranmayee et al. (2016) proposed a brain MRI classification method using a SVM. In the data processing stage, a median adaptive filter was used to remove noise, and then the watershed method, fuzzy clustering method, and threshold method were used to segment MRI brain image. The kernel SVM was used as the classifier.

The dictionary learning method is widely used to solve various problems of computer vision and image analysis (Gu et al., 2020; Ni et al., 2020). Dictionary learning aims to find a suitable dictionary for the input data and transform it into a sparse representation, so as to mine the useful features of the data, simplify the learning task and reduce the complexity of the model. A kernel sparse representation was developed in Chen et al. (2017). It contained three key steps for multi-label brain tumor segmentation: component analysis-split for dictionary learning initialization, kernel dictionary learning and kernel sparse coding, and graph-cut method for image segmentation. A system combining an adaptive type-2 fuzzy system and dictionary learning was proposed in Ghasemi et al. (2020), in which the sparse coding step and dictionary learning step were executed alternately, and the fuzzy membership functions in the type-2 fuzzy system were used to represent model uncertainty and improve sparse representation. A learning method combining discriminate sub-dictionary and projective dictionary pair learning was developed for classifying proton magnetic resonance spectroscopy of brain gliomas tumor (Adebileje et al., 2017).

AI mainly uses intelligent methods to extract brain image features, which requires a large number of labeled data sets to understand the potential connections in the data. But in the field of medicine, because of the confidentiality and professionalism of patient information, medical data need to be marked by professional clinicians, and data collection is complex and expensive. Lack of labeled trainable data is one of the bottlenecks that restrict the development of medical image analysis. In addition, traditional AI methods require training data and test data to be independent and identically distributed. Transfer learning relaxes this restriction on training data and test data (Ni et al., 2018b; Jiang et al., 2020; Jiang et al., 2021). It can apply the knowledge or patterns learned from a related domain (source domain) to another target domain, and utilize the information shared by source domain samples and target domains, then finally build a model to adapt to the target domain.

To solve this problem, this paper focuses on solving the distribution differences between source and target domains. Through the feature mapping of source and the target domain samples, the source domain knowledge can be transferred to

target domain learning. Because dictionary learning can exploit the essential characteristics of the data, this paper uses Multi-layer dictionary learning (MDL) in transfer learning to exploit the shared knowledge between source and target domains. MDL first obtains the dictionary and sparse features of the first layer on the original samples, then obtains the dictionary and sparse features of the second layer based on the obtained sparse features of the first layer, and learns the dictionary and sparse features in turn to finally obtain the deep dictionary and sparse features. Finally, the new test data can be encoded by the multi-layer dictionary and the final classification results can be obtained. According to the difference of domain and task, transfer learning is divided into feature transfer, sample transfer and parameter transfer. In this paper, the target and source domain are images, and the task is to train the image, extract features, and realize the classification of different types of images, so this paper belongs to the parameter transfer mode. The advantages of this algorithm are as follows: (1) based on multi-layer learning, multi-layer dictionaries are obtained, and the discriminability of sparse representation coefficients can be enhanced in layer by layer dictionary learning; (2) through multi-layer shared dictionary learning, the sample reconstructions of source and target domains are constrained layer by layer, so as to minimize the error of sample reconstruction both in source and target domains; (3) by utilizing the label information, Laplacian regularization term is introduced, and the sparse coding of samples in the same class is as close as possible, while the sparse coding of samples in different classes is as different as possible. At the same time, in the last layer of the proposed model, the classification error term is introduced in the last layer of MDL to improve the discriminative performance of the model; (4) The recognition experiments on brain MRI image datasets REMBRANDT (Clark et al., 2013) and Figshare (Cheng et al., 2016) show that the proposed model performs satisfactory classification performance in terms of accuracy, precision, F1-score, and recall.

The rest of the paper is organized as follows: the related work is introduced in section "Backgrounds." The proposed method is given in section "Proposed Method", and experiments are performed in section "Experiment." Finally, conclusion and future work are summarized in section "Conclusion."

BACKGROUNDS

Dictionary Learning

Dictionary learning methods can basically be divided into unsupervised dictionary learning and supervised dictionary learning. The unsupervised dictionary learning does not make use of sample label information. The supervised dictionary learning makes use of sample label information and pays more attention to the discriminative ability of sparse representation coefficients.

KSVD (Jiang et al., 2013) is a famous supervised dictionary learning algorithm. KSVD introduces the classification error of a linear classifier into the objective function, while learning the representation and classification ability of the dictionary. The objective function of K-SVD is

$$< \mathbf{D},\mathbf{W},\mathbf{Z} > = min||\mathbf{X}\text{-}\mathbf{DZ}||_2^2 + \gamma||\mathbf{H}\text{-}\mathbf{WZ}||_2^2 + \beta||\mathbf{W}||_2^2 \quad s.t.\ ||\mathbf{Z}||_0 \le T \tag{1}$$

where $\mathbf{Z}$ is the sparse representation coefficient, $\mathbf{W}$ is the parameter of the linear classifier, $\mathbf{H}$ is the label vector of the training data. To solve Eq. (1), the first two of these terms are combined and Eq. (1) is rewritten as

$$< \mathbf{D},\mathbf{W},\mathbf{Z} >= \min || \begin{pmatrix} \mathbf{X} \\ \sqrt{\gamma}\mathbf{H} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\gamma}\mathbf{W} \end{pmatrix} \mathbf{Z}||_2^2 + \beta||\mathbf{W}||_2^2 \quad s.t.\ ||\mathbf{Z}||_0 \le T \tag{2}$$

Eq. (2) can be solved by using an iterative strategy. When $\mathbf{W}$ is fixed, the problem of $<\mathbf{D},\mathbf{Z}>$ represents the same formulation as K-SVD, and it can therefore be solved using the K-SVD. When $\mathbf{D}$ and $\mathbf{Z}$ are fixed, Eq. (2) is a simple linear problem that can be solved by linear methods.

Multi-Layer Dictionary Learning

With the development of deep learning, researchers have found that the deeper the structure of a neural network, the better and more accurate the representation. MDL (also known as deep dictionary learning) refers to the idea of deep learning, and applies "deep structure" to layer-by-layer dictionary learning (Song et al., 2019; Gu et al., 2020). The dictionary and sparse representation obtained by the traditional single-layer dictionary learning method are shallow, which is not conducive to the task of recognition and classification when the data dimension is too high or the number of samples is too large. Singhal et al. (2017) proposed a deep dictionary learning model, which used the idea of deep learning to learn the multi-level dictionary and the deep features of the original samples. As an example, the two-layer dictionary learning is illustrated in **Figure 1**. $\mathbf{D}_1$ and $\mathbf{D}_2$ are dictionaries learned in the first and second layer. $\mathbf{Z}_2$ is the sparse coefficient learned in the second layer. The sample $\mathbf{X}$ can be represented as $X = \mathbf{D}_1\mathbf{Z}_1 = \mathbf{D}_1\mathbf{D}_2\mathbf{Z}_2$, where the sparse coding learning in the first layer $\mathbf{Z}_1 = \mathbf{D}_2\mathbf{Z}_2$. Specifically, the first layer is solved as a single layer of dictionary learning to

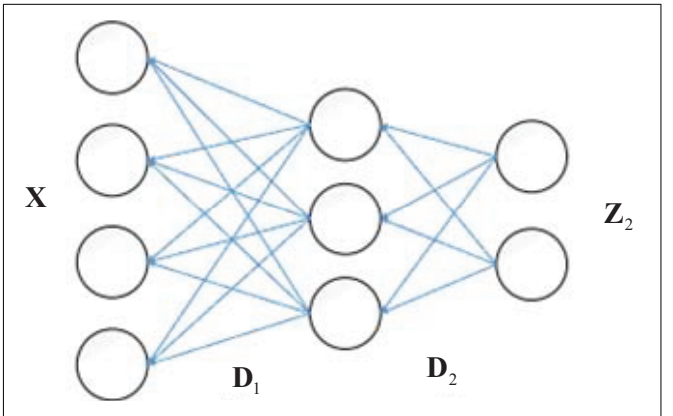


FIGURE 1 | The schematic diagram of two-layer dictionary learning.

obtain feature $\mathbf{Z}_1$ on dictionary $\mathbf{D}_1$, and $\mathbf{Z}_1$ is then used as input to the second layer, which is also solved as a single layer of dictionary learning to obtain feature $\mathbf{Z}_2$. The new test data can be encoded by the learned $\mathbf{D}_1$ and $\mathbf{D}_2$. In this way, after completing the L-layer dictionary learning, the final dictionary and sparse representations are obtained as $\mathbf{D}_L$ and $\mathbf{Z}_L$. In this case, the sample $\mathbf{X}$ can be represented as

$$\mathbf{X} = \mathbf{D}_1(\mathbf{D}_2(...(\mathbf{D}_L\mathbf{Z}_L))) \tag{3}$$

Then the dictionaries in L-layers and the sparse coding can be solved by

$$\min_{\mathbf{D}_1,\mathbf{D}_2,...,\mathbf{D}_L,\mathbf{Z}_L} \left|\left|\mathbf{X}\text{-}\mathbf{D}_1(\mathbf{D}_2(...(\mathbf{D}_L\mathbf{Z}_L)))\right|\right|_F^2 + \left|\left|\mathbf{Z}_L\right|\right|_1 \tag{4}$$

PROPOSED METHOD

Objective Function

We assume that there is a corresponding association between source and target domains in transfer learning. From this point, based on the framework of MDL, we try to learn the common shared dictionary between source and target domains to exploit the shared knowledge among different related domains. At the same time, by making full use of the label information of the samples, the classification error term is introduced in the last layer of the multi-layer dictionary, which makes the sparse representation of the target domain more discriminative. According to this idea, we propose a transfer model based on supervised multi-layer dictionary learning (TSMDL), and its objective function is

$$\begin{aligned}
\min_{\mathbf{Z},\mathbf{w},b} & \left\|\mathbf{X}_s - \mathbf{D}_1\mathbf{Z}_1^s\right\|_F^2 + \left|\left|\mathbf{X}_t - \mathbf{D}_1\mathbf{Z}_1^t\right|\right|_F^2 + \alpha_1 \sum_{i,j} Q_{1,i,j}^{C,s}\left|\left|\mathbf{z}_{1,i}^s - \mathbf{z}_{1,j}^s\right|\right|_F^2 \\
& -\beta_1 \sum_{i,j} Q_{1,i,j}^{M,s}\left|\left|\mathbf{z}_{1,i}^s - \mathbf{z}_{1,j}^s\right|\right|_F^2 + \alpha_1 \sum_{i,j} \mathbf{Q}_{1,i,j}^{C,t}\left|\left|\mathbf{z}_{1,i}^t - \mathbf{z}_{1,j}^t\right|\right|_F^2 \\
& -\beta_1 \sum_{i,j} Q_{1,i,j}^{M,t}\left|\left|\mathbf{z}_{1,i}^t - \mathbf{z}_{1,j}^t\right|\right|_F^2 + ... + \left|\left|\mathbf{Z}_{L-1}^s - \mathbf{D}_L\mathbf{A}_L^s\right|\right|_F^2 \\
& +\left|\left|\mathbf{z}_{L-1}^t - \mathbf{D}_L\mathbf{A}_L^t\right|\right|_F^2 + \alpha_L \sum_{i,j} Q_{L,i,j}^{C,s}\left|\left|\mathbf{z}_{L,i}^s - \mathbf{z}_{L,j}^s\right|\right|_F^2 \\
& -\beta_L \sum_{i,j} Q_{L,i,j}^{M,s}\left|\left|\mathbf{z}_{L,i}^s - \mathbf{z}_{L,j}^s\right|\right|_F^2 + \alpha_L \sum_{i,j} \mathbf{Q}_{L,i,j}^{C,t}\left|\left|\mathbf{z}_{L,i}^t - \mathbf{z}_{L,j}^t\right|\right|_F^2 \\
& -\beta_L \sum_{i,j} \mathbf{Q}_{L,i,j}^{M,t}\left|\left|\mathbf{z}_{L,i}^t - \mathbf{z}_{L,j}^t\right|\right|_F^2 + \lambda \sum_{c=1}^{C_s} f(\mathbf{z}_L^s, \mathbf{y}_c^s, \mathbf{w}_c^s, \mathbf{b}_c^s) \\
& +\lambda \sum_{c=1}^{C_t} f(\mathbf{z}_L^t, \mathbf{y}_c^t, \mathbf{w}_c^t, \mathbf{b}_c^t) + \lambda_l \sum_{l=1}^{L} \left|\left|\mathbf{z}_l\right|\right|_F^2, \\
\text{s.t. } & \left|\left|d_l^j\right|\right|_2^2 = 1, l = 1, 2, ..., L, \quad j = 1, 2, ..., K_l
\end{aligned} \tag{5}$$

where K_l is the size of dictionary in the lth layer. $\mathbf{Z}_l^s$ and $\mathbf{Z}_l^s$ are the spares coding matrixes in the l-th layer. $\mathbf{Q}_l^{C,(\cdot)}$ and $\mathbf{Q}_l^{M,(\cdot)}$ are the matrixes of weights for data samples of the same class and data samples of different classes in the l-th layer. $Q_{l,ij}^{C,(\cdot)}$ and $Q_{l,ij}^{M,(\cdot)}$ can be defined as

$$Q_{l,ij}^{C,(\cdot)} = \begin{cases} \frac{1}{|C|}, & (\mathbf{z}_{l,i}^{(\cdot)}, \mathbf{z}_{l,j}^{(\cdot)}) \text{ belonging to the same class} \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

$$Q_{l,ij}^{M,(\cdot)} = \begin{cases} \frac{1}{|M|}, & (\mathbf{z}_{l,i}^{(\cdot)}, \mathbf{z}_{l,j}^{(\cdot)}) \text{ belonging to different classes} \\ 0, \text{ otherwise} \end{cases} \tag{7}$$

where $(\cdot)$ means s or t.

We explain the above Eq. (5) as follows:

1. The first two terms $\left|\left|\mathbf{X}_s - \mathbf{D}_1\mathbf{Z}_1^s\right|\right|_F^2$ and $\left|\left|\mathbf{X}_t - \mathbf{D}_1\mathbf{Z}_1^t\right|\right|_F^2$ are the reconstruction error terms of source domain and target domain data in the first layer of the learning framework.
2. The third and fourth terms $\sum_{i,j} \mathbf{Q}_{1,i,j}^{C,s}\left|\left|\mathbf{z}_{1,i}^s - \mathbf{z}_{1,j}^s\right|\right|_F^2$ and $\sum_{i,j} \mathbf{Q}_{1,i,j}^{M,s}\left|\left|\mathbf{z}_{1,i}^s - \mathbf{z}_{1,j}^s\right|\right|_F^2$ are the Laplacian regularization terms of the source domain in the first layer, which, respectively, constrain the dictionary codes of the same class in the source domain to be as close as possible, and the dictionary codes of different classes to be as different as possible. The fifth and sixth terms $\sum_{i,j} \mathbf{Q}_{1,i,j}^{C,t}\left|\left|\mathbf{z}_{1,i}^t - \mathbf{z}_{1,j}^t\right|\right|_F^2$ and $\sum_{i,j} \mathbf{Q}_{1,i,j}^{M,t}\left|\left|\mathbf{z}_{1,i}^t - \mathbf{z}_{1,j}^t\right|\right|_F^2$ are the Laplacian regularization terms of the target domain in the first layer. Similarly to the third and fourth terms, their goal is to, respectively, constrain the dictionary codes of the same class in the target domain to be as close as possible, and the dictionary codes of different classes to be as different as possible.
3. Following the generation rules for the first six terms, the corresponding reconstruction error terms and Laplacian regularization terms for the source and target domains are constructed for layers 2 to L.
4. $\sum_{c=1}^{C_s} f(\mathbf{Z}_L^s, \mathbf{y}_c^s, \mathbf{w}_c^s, \mathbf{b}_c^s)$ and $\sum_{c=1}^{C_t} f(\mathbf{Z}_L^t, \mathbf{y}_c^t, \mathbf{w}_c^t, \mathbf{b}_c^t)$ are classification error terms for the last layer of the source domain and target domain, respectively. Its goal is to improve the discriminative ability of the model. In this study, we use SVM multi-class classifier. The parameters $w_c^{(\cdot)}$ and $b_c^{(\cdot)}$ are hyperplane parameters in the SVM.

Define Laplacian matrix in the same class $\mathbf{P}_l^{C,(\cdot)}$ as $\mathbf{P}_l^{C,(\cdot)} = \tilde{\mathbf{Q}}_l^{C,(\cdot)} - \mathbf{Q}_l^{C,(\cdot)}$, where $\tilde{\mathbf{Q}}_{l,ii}^{C,(\cdot)} = \sum_j \mathbf{Q}_{l,ij}^{C,(\cdot)}$, Laplacian matrix

TABLE 1 | The basic information of transfer learning tasks in the experiment.

Tasks	Source domain		Target domain	
T1	Normal images from the REMBRANDT dataset	Tumors images from the Figshare dataset	Normal images from the REMBRANDT dataset	Tumors images from the REMBRANDT dataset
T2	Normal images from the REMBRANDT dataset	Tumors images from the REMBRANDT dataset	Normal images from the REMBRANDT dataset	Tumors images from the Figshare dataset

in the different classes $\mathbf{P}_l^{M,(\cdot)}$ as $\mathbf{P}_l^{M,(\cdot)} = \tilde{\mathbf{Q}}_l^{M,(\cdot)} - \mathbf{Q}_l^{M,(\cdot)}$, where $\tilde{\mathbf{Q}}_{l,ii}^{M,(\cdot)} = \sum_j Q_{l,ij}^{M,(\cdot)}$. Let $\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_t]$, $\mathbf{Z} = [\mathbf{Z}_s, \mathbf{Z}_t]$, $\mathbf{P}_l^C = \begin{bmatrix} \mathbf{P}_l^{C,s} & 0 \\ 0 & \mathbf{P}_l^{C,t} \end{bmatrix}$, $\mathbf{P}_l^M = \begin{bmatrix} \mathbf{P}_l^{M,s} & 0 \\ 0 & \mathbf{P}_l^{M,t} \end{bmatrix}$, Eq. (5) can be written as

$$\begin{aligned} \min_{\mathbf{D}_l,\mathbf{P}_l^C,\mathbf{P}_l^M,Z_l,w,b} & \left|\mathbf{X}\text{-}\mathbf{D}_1\mathbf{Z}_1\right|_F^2 + Tr(\mathbf{Z}_1(\alpha_1\mathbf{P}_1^C + \lambda_1\mathbf{I})\mathbf{Z}_1^T) \\ & -\beta_1 Tr(\mathbf{Z}_1\mathbf{P}_1^M\mathbf{Z}_1^T) + ...+ \\ & \left|\left|\mathbf{Z}_{L-1} - \mathbf{D}_L\mathbf{Z}_L\right|\right|_F^2 + Tr(\mathbf{Z}_L(\alpha_L\mathbf{P}_L^C + \lambda_L\mathbf{I})\mathbf{Z}_L^T) \\ & -\beta_L Tr(\mathbf{Z}_L\mathbf{P}_L^M\mathbf{Z}_L^T) + \lambda\textstyle\sum_{c=1}^{\mathbf{C}} f(\mathbf{Z}_L, \mathbf{y}_c, \mathbf{w}_c, \mathbf{b}_c), \\ \textbf{s.t.}\ & \left|\left|d_l^j\right|\right|_2^2 = 1,\ l = 1, 2, ..., L,\quad j = 1, 2, ..., K_l, \end{aligned} \tag{8}$$

Again, we simplify the function above and obtain that

$$\begin{aligned} \min_{\mathbf{D}_l,\mathbf{P}_l^C,\mathbf{P}_l^M,\mathbf{Z}_l,w,b} & \left|\left|\mathbf{X}\text{-}\mathbf{D}_1\mathbf{Z}_1\right|\right|_F^2 + Tr(\mathbf{Z}_1(\alpha_1\mathbf{P}_1^C - \beta_1\mathbf{P}_1^M + \lambda_1\mathbf{I}) \\ & \mathbf{Z}_1^T + ...+ \\ & \left|\left|\mathbf{Z}_{L-1} - \mathbf{D}_L\mathbf{Z}_L\right|\right|_F^2 + Tr(\mathbf{Z}_L(\alpha_L\mathbf{P}_L^C - \beta_L\mathbf{P}_L^M \\ & +\lambda_L\mathbf{I})\mathbf{Z}_L^T) + \lambda\textstyle\sum_{c=1}^{C} f(\mathbf{Z}_L, \mathbf{y}_c, \mathbf{w}_c, \mathbf{b}_c), \\ \textbf{s.t.}\ & \left|\left|d_l^j\right|\right|_2^2 = 1,\ l = 1, 2, ..., L,\quad j = 1, 2, ..., K_l, \end{aligned} \tag{9}$$

Optimization

We use the alternating optimization approach to solve Eq. (9). The parameters to be solved include $\mathbf{D}_1$, $\mathbf{P}_1^C$, $\mathbf{P}_1^M$, $\mathbf{Z}_1$,. . ., $\mathbf{D}_L$, $\mathbf{P}_L^C$, $\mathbf{P}_L^M$, $\mathbf{Z}_L$, $\boldsymbol{w}$ and b. In the following, we divide the solution of these variables into three parts.

a. Update parameters $\mathbf{D}_1$, $\mathbf{P}_1^C$, $\mathbf{P}_1^M$, $\mathbf{Z}_1$,. . ., $\mathbf{D}_L$, $\mathbf{P}_L^C$ and $\mathbf{P}_L^M$

First, we update parameters $\mathbf{D}_1$, $\mathbf{P}_1^C$, $\mathbf{P}_1^M$ and $\mathbf{Z}_1$in the first layer. When fixed the other parameters, the objective function of TSMDL is

$$\begin{aligned} \min_{\mathbf{D}_1,\mathbf{P}_1^C,\mathbf{P}_1^M,\mathbf{Z}_1} & \left|\left|\mathbf{X}\text{-}\mathbf{D}_1\mathbf{Z}_1\right|\right|_F^2 + Tr(\mathbf{Z}_1(\alpha_1\mathbf{P}_1^C - \beta_1\mathbf{P}_1^M + \lambda_1\mathbf{I})\mathbf{Z}_1^T \\ & +\left|\left|\mathbf{Z}_1 - \mathbf{D}_2\mathbf{Z}_2\right|\right|_F^2, \\ \textbf{s.t.}\ & \left|\left|d_1^j\right|\right|_2^2 = 1 \end{aligned} \tag{10}$$

Further, the parameters except for $\mathbf{D}_1$ are fixed, the optimization problem can be written as

$$\begin{aligned} \min_{\mathbf{D}_1,\mathbf{P}_1^C,\mathbf{P}_1^M,\mathbf{Z}_1} & \left|\left|\mathbf{X}\text{-}\mathbf{D}_1\mathbf{Z}_1\right|\right|_F^2, \\ \textbf{s.t.}\ & \left|\left|d_1^j\right|\right|_2^2 = 1 \end{aligned} \tag{11}$$

Following (Boyd et al., 2011), the optimal value of $\mathbf{D}_1$ can be computed by an alternating direction method of multipliers. Then the Laplacian matrixes $\mathbf{P}_1^C$ and $\mathbf{P}_1^M$ can be computed according to Eqs.(6, 7).

The optimal value of $\mathbf{Z}_1$ can be obtained by taking the derivation of Eq.(8) as the following formulation, i.e.,

$$\mathbf{Z}_1 = (\mathbf{D}_1^T\mathbf{D}_1 + \alpha_1\mathbf{P}_1 - \beta_1\mathbf{P}_1^M + (\lambda_1 + 1)\mathbf{I})^{-1}(\mathbf{D}_1^T\mathbf{X} + \mathbf{D}_2\mathbf{Z}_2) \tag{12}$$

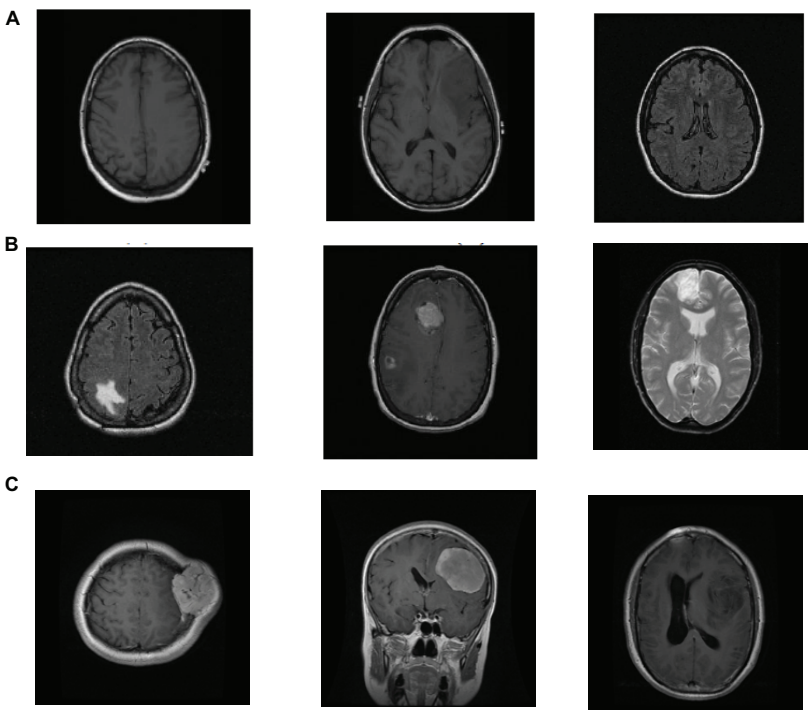


FIGURE 2 | Example samples used in the experiment, **(A)** normal images in the REMBRANDT dataset, **(B)** tumor images in the REMBRANDT dataset, **(C)** tumor images in the Figshare dataset.

For $2 \leq l \leq L$, when fixing the other parameters, the objective function of $\mathbf{D}_l$ is

$$\begin{aligned} &\min_{\mathbf{D}_l} \left|\left|\mathbf{Z}_{l-1} - \mathbf{D}_l\mathbf{Z}_l\right|\right|_F^2, \\ &\textbf{s.t.} \ \left|\left|d_2^j\right|\right|_2^2 = 1, j = 1, 2, ..., K_L \end{aligned} \tag{13}$$

After obtaining the $\mathbf{D}_l$, the optimal value of $\mathbf{Z}_l (2 \leq l \leq L-1)$ can be obtained by,

$$\mathbf{Z}_l = (\mathbf{D}_l^T\mathbf{D}_l + \alpha_l\mathbf{P}_l^C - \beta_l\mathbf{P}_l^M + (\lambda_l + 1)\mathbf{I})^{-1}(\mathbf{D}_1^T\mathbf{X} + \mathbf{D}_{l+1}\mathbf{Z}_{l+1}) \tag{14}$$

b. Update parameter $\mathbf{Z}_L$:

When the other parameters are fixed, the objective function of TSMDL related to $\mathbf{Z}_L$ is

$$\begin{aligned} \min_{\mathbf{Z}_L} \quad &\left|\left|\mathbf{Z}_{L-1} - \mathbf{D}_L\mathbf{Z}_L\right|\right| + Tr(\mathbf{Z}_L(\alpha_L\mathbf{P}_L^C - \beta_L\mathbf{P}_L^M + \lambda_L\mathbf{I})\mathbf{Z}_L^T) \\ &+\lambda\sum_{c=1}^{C} f(\mathbf{Z}_L, \mathbf{y}_c, \mathbf{w}_c, \mathbf{b}_c) \end{aligned} \tag{15}$$

Let $\mathbf{z}_L^i (i = 1, 2, ..., N)$ be the ith column of $\mathbf{Z}_L$. We rewrite Eq. (15) related to $\mathbf{z}_L^i$ as

$$\begin{aligned} \min_{\mathbf{Z}_L} \quad &\left|\left|\mathbf{z}_{L-1}^i - \mathbf{D}_L\mathbf{Z}_L\right|\right| + Tr(\mathbf{z}_L^i(\alpha_L\mathbf{P}_L^C - \beta_L\mathbf{P}_L^M + \lambda_L\mathbf{I})\mathbf{z}_L^{iT}) \\ &+\lambda\sum_{c=1}^{C} f(\mathbf{z}_L^i, \mathbf{y}_c^i, \mathbf{w}_c, \mathbf{b}_c) \end{aligned} \tag{16}$$

In this study, we use standard L1-SVM for term $f(\mathbf{z}_L^i, \mathbf{y}_c^i, \mathbf{w}_c, \mathbf{b}_c)$, thus we can set $\mathbf{y}_c^i = 1$ if class label $\mathbf{y}_c^i = c$ and otherwise $y_c^i = -1$. In this case, the optimal value of $\mathbf{z}_L^i$ can be computed by a least square problem.

c. Update parameters $\mathbf{w}$ and b

When the other parameters are fixed, the objective function of TSMDL related to $\mathbf{w}$ and b is

$$\min_{\mathbf{w}_c, \mathbf{b}_c} \sum_{c=1}^{C} f(\mathbf{Z}_L, \mathbf{y}_c, \mathbf{w}_c, b_c), \tag{17}$$

Obviously, Eq. (17) can be solved by various SVM solvers.

We show the optimization procedure of TSMDL in algorithm 1.

Input: Training data matrix $\mathbf{X}$, parameters α_l, β_l and λ_l, $\forall\, l$

1: Initialize $\mathbf{D}$ using K-SVD algorithm on each class, initialize $\mathbf{P}$ using principal component analysis (PCA) algorithm;

2: **while** not converged **do**

3: Compute $\mathbf{D}_l (1 \leq l \leq L)$ by Eq. (11);

4: Compute the Laplacian matrixes $\mathbf{P}_l^C$ and $\mathbf{P}_l^M$ $(1 \leq l \leq L)$ by Eqs. (6, 7);

5: Compute $\mathbf{Z}_l (1 \leq l \leq L-1)$ by Eqs. (12–14);

6: Compute $\mathbf{z}_L^i$ by Eq. (16);

7: Compute $\mathbf{w}$ and b by Eq. (17);

8: **end while**

Output: $\mathbf{D}_1, \mathbf{P}_1^C, \mathbf{P}_1^M, \mathbf{Z}_1, \ldots, \mathbf{D}_L, \mathbf{P}_L^C, \mathbf{P}_L^M, \mathbf{Z}_L$, $\mathbf{w}$ and b.

Learning a Classifier

We compute $\mathbf{\Theta_l} = (\mathbf{D}_l^T\mathbf{D}_l + \alpha\mathbf{P}_l^C - \beta\mathbf{P}_l^M + \lambda_l\mathbf{I})^{-1}\mathbf{D}_l^T (l = 1, 2, ..., L)$. The test sample $\mathbf{X}_{\text{new}}$, we compute its sparse coding as $z_{\text{new}} = \mathbf{\Theta}_1 ... \mathbf{\Theta}_L \mathbf{x}_{\text{new}}$. Finally, we can use the following formulation to predict the class label of $\mathbf{x}_{\text{new}}$

$$\mathbf{y} = \arg\max_c \ \mathbf{w}_c^T\mathbf{z}_{\text{new}} + b_c, \ \forall c \tag{18}$$

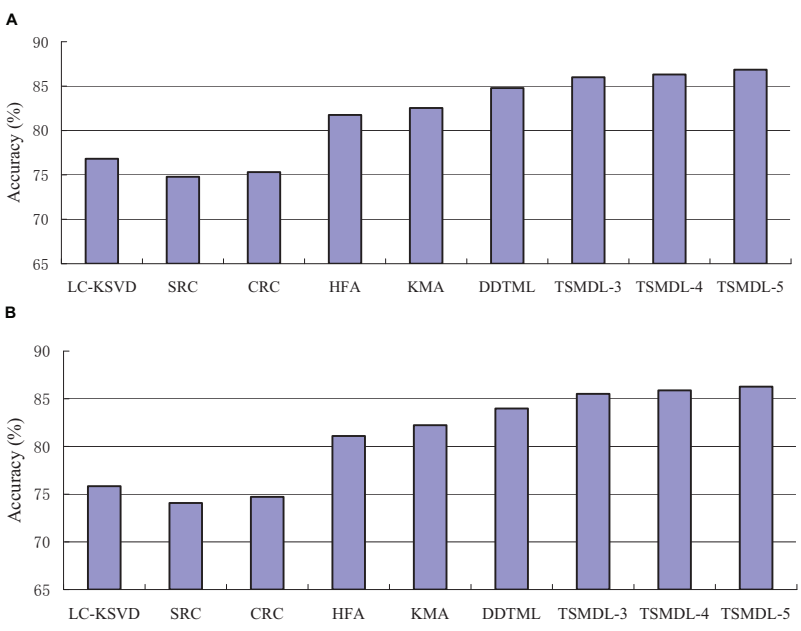


FIGURE 3 | Accuracy comparison results on, **(A)** T1 task, **(B)** T2 task.

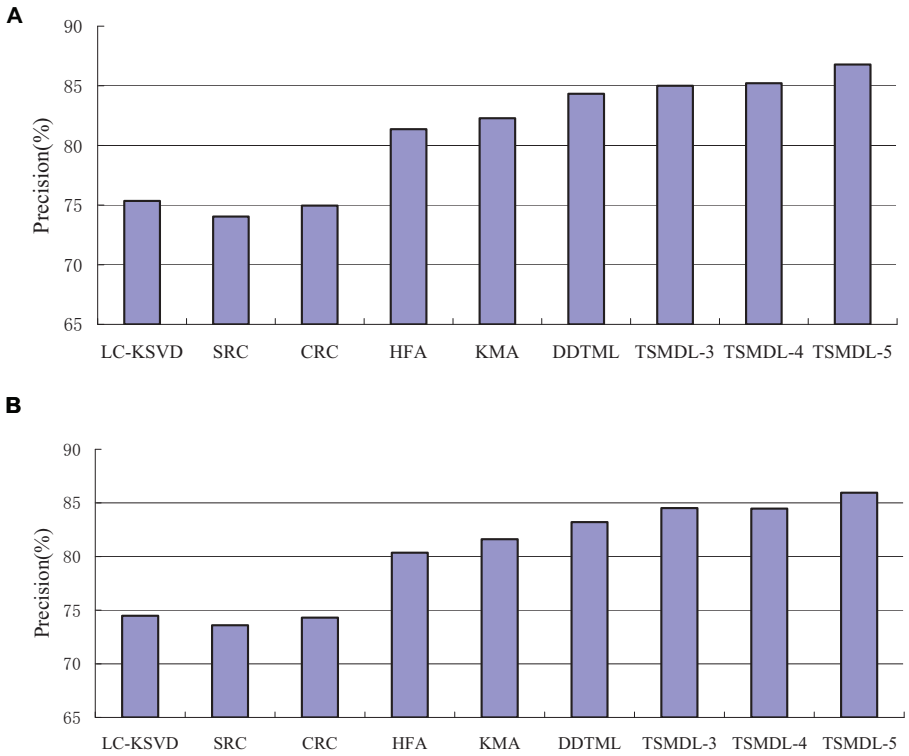


FIGURE 4 | Precision comparison results on tasks, **(A)** T1, **(B)** T2.

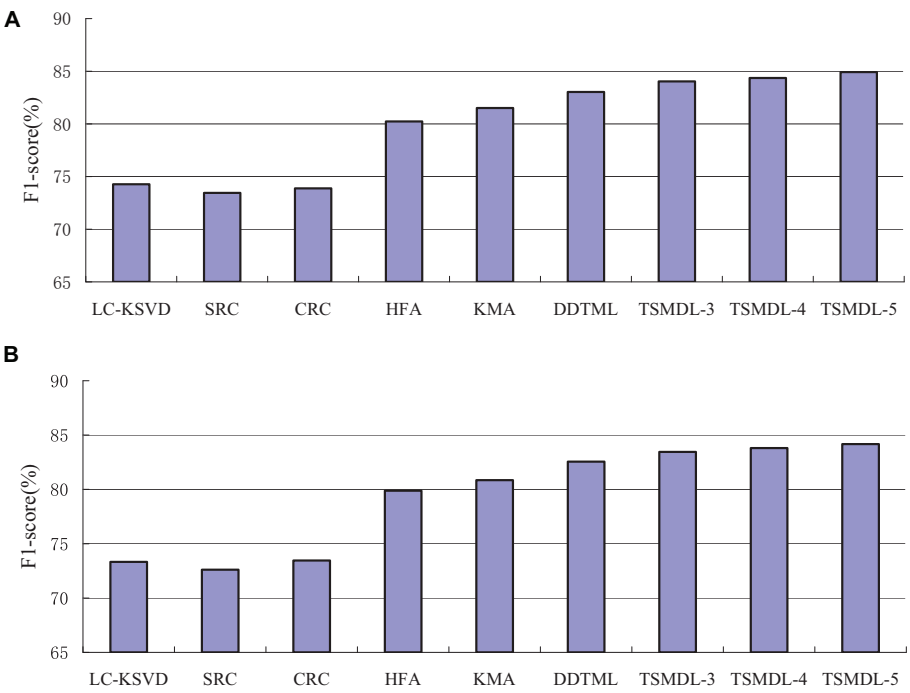


FIGURE 5 | F-Score comparison results on tasks, **(A)** T1, **(B)** T2.

EXPERIMENT

Experiment Settings

The datasets used in the study are taken from (Clark et al., 2013) and Figshare (Cheng et al., 2016) datasets. Figshare dataset is collected from two hospitals in China between 2005 and 2010. It contains a vast number of MRI images from 233 patients with brain tumors, including meningiomas, pituitary and gliomas. All images are digitized at a resolution of 512 × 512 pixels. REMBRANDT dataset contains a vast number of MRI images collected from 130 brain tumor patients. The patients' ages ranged from 15 to 89 years, with a mean of 47.5 years. The MRI images included astrocytomas from 47 patients; oligodendrogliomas from 21 patients; GBMs from 40 patients and the truth of the tumor in the remaining patients is unknown. All images are digitized at a resolution of 256 × 256 pixels. Some slices in the REMBRANDT dataset, where the tumor lesions are found, are considered normal samples. In the experiment, we design two transfer learning tasks, and show their information in **Table 1**. **Figure 2** shows sample images in the REMBRANDT and Figshare datasets. The main objective of two tasks is to classify the brain MRI images into normal and tumor classes. In task T1, we randomly select 200 normal images and 200 tumor images from the REMBRANDT dataset as source domain, and randomly select 200 normal images from the REMBRANDT dataset and 200 tumor images from Figshare dataset as target domain. There are no duplicate images in source and target domains. We use all images in the source domain and 10% images in the target domain as training data, and use the rest of the images in the target domain as testing data. We use wavelet transform wavelets and gray level co-occurrence matrix (GLCM) method for feature extraction (Mohankumar, 2016). Each image is extracted onto a 540 dimensional vector.

In the experiment, we compare our model with LC-KSVD (Jiang et al., 2013), SRC (Wright et al., 2009), CRC (Zhang et al., 2011), HFA (Long et al., 2013), KMA (Tuia and Camps-Valls, 2016), and DDTML (Ni et al., 2018a). Following the authors, all parameters in comparative methods are set in their default settings. The parameters β, λ_1, and λ in TSMDL are set in the grid {0.01, 0.05, 0.1,...,2}. The number of layers is set in {3, 4, 5}, and the TSMDL model is accordingly named as TSMDL-3, TSMDL-4, and TSMDL-5, respectively. The sizes of dictionaries are 500, 450, 400, 350, and 300 corresponding to layer 1 to layer 5, respectively. In order to ensure the stability and effectiveness of the experimental results, for the proposed model and other comparative experimental methods, we run each task 10 times. All the methods are implemented in MATLAB, and the environment that we used in the experiments is a computer with Intel Core i5-3317U 1.70 GHz CPU, 16 GB RAM.

Experiment Results

In this subsection, we present the effect of TSMDL on T1 and T2 tasks. We summarize the performance of all comparative methods in terms of accuracy, precision, F1-score, and recall. The experiment results are shown in **Figures 3–6**, respectively. According to **Figures 3–6**, we can draw the following results:

In terms of accuracy, precision, F1-score, and recall, the proposed TSMDL achieves the best results. In addition,

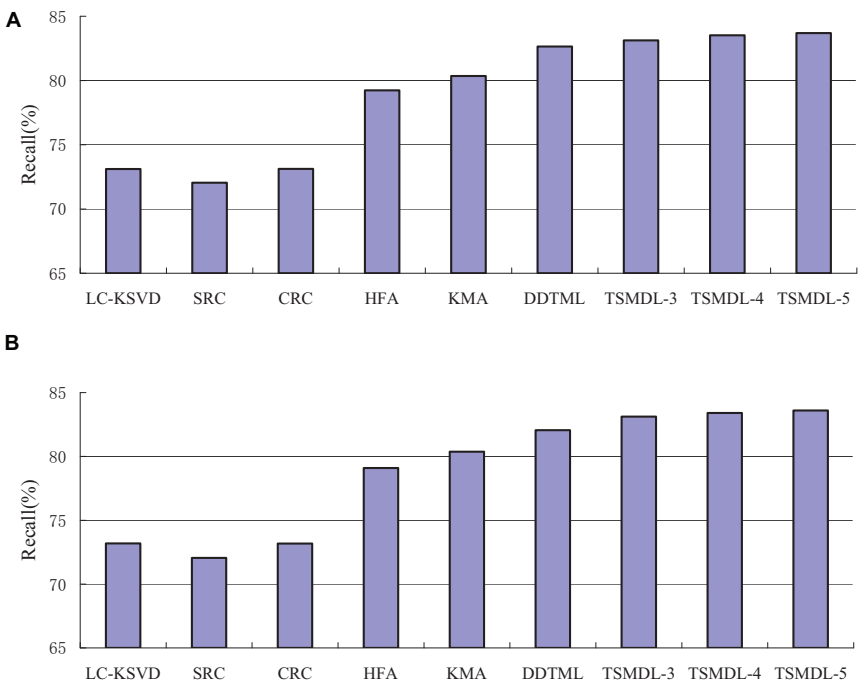


FIGURE 6 | Recall comparison results on tasks, **(A)** T1, **(B)** T2

the performance of TSMDL-5 is better than TSMDL-3 and TSMDL-4. It is indicated that the multi-layer framework of dictionary learning can exploit the instinct structure of data samples and can build a relationship between source and target domains. Thus, TSMDL is suitable for the application of brain tumor MRI image recognition.

In the experiments, except for the LC-KSVD, SRC, and CRC algorithms, all other algorithms are transfer learning-based classification methods, which show that transfer learning strategy is helpful for brain tumor MRI image classification in the target domain. The classification knowledge in the source domain can be effectively transferred to the target domain to help the target domain achieve better classification results.

The proposed TSMDL in this paper is obviously superior to other transfer learning methods, which shows that multiple layer transfer learning dictionary learning can truly restore the brain MRI images of source and target domains, and reduce the distribution difference between domains. Thus, it can strengthen the domain adoption between source and target domains in the sparse representation space. The reason is that TSMDL is based on MDL; it can learn a more complex and accurate dictionary to represent the original data, and obtain more discriminative representation coefficients. In addition, TSMDL is a supervised learning model, in which the label information can be exploited, so TSMDL can obtain higher discrimination performance.

CONCLUSION

With the popularity of MRI equipment, a large number of new MRI brain images emerge, but obtaining labeled data is very time-consuming and expensive. Therefore, the goal of this paper is to use a large number of labeled data from the source domain to learn a classifier with strong generalization ability, and to classify the target domain with only a small number of labeling samples. Therefore, based on the MDL framework, we learn the common dictionary on each layer of the network to minimize the sample reconstruction error of the constrained source domain and target domain. At the same time, the Laplacian regularization term is introduced in each layer of the network to make the sparse coding of similar samples as close as possible, while the sparse coding of different classes of samples is as different as possible. The experimental results on brain MRI image datasets REMBRANDT and Figshare show that our model achieves the state-of-the-art methods. Future works will include studying the effect of using unlabeled samples in the target domain while training, and other relevant problems like large-scale and online adaptation of dictionaries.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The download URLs of the REMBRANDT and Figshare datasets are, respectively, [https://wiki.cancerimagingarchive.net/display/Public/REMBRANDT] and [https://figshare.com/articles/dataset/brain_tumor_dataset/1512427].

AUTHOR CONTRIBUTIONS

YG developed the theoretical framework and model in this work and drafted the manuscript. YG and KL implemented the algorithm and performed experiments and result analysis. Both contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grants 61772241 and 61873321, in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

REFERENCES

Adebileje, S. A., Ghasemi, K., Aiyelabegan, H. T., and Rad, H. S. (2017). Accurate classification of brain gliomas by discriminate dictionary learning based on projective dictionary pair learning of proton magnetic resonance spectra. *Magn. Reson. Chem.* 55, 318–322. doi: 10.1002/mrc.4532

Amin, J., Sharif, M., Yasmin, M., and Fernandes, S. L. (2017). A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognit. Lett.* 139, 118–127. doi: 10.1016/j.patrec.2017.10.036

Bahadure, N. B., Ray, A. K., and Thethi, H. P. (2017). Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. *Int. J. Biomed. Imaging* 2017:9749108. doi: 10.1155/2017/9749108

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends.* 3, 1–128. doi: 10.1561/2200000016

Chen, X., Nguyen, B. P., Chui, C. K., and Ong, S. H. (2017). An automatic framework for multi-label brain tumor segmentation based on kernel sparse representation. *Acta Polytech. Hung.* 14, 25–43. doi: 10.12700/APH.14.1.2017.1.3

Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., et al. (2016). Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PLoS One* 11:e0157112. doi: 10.1371/journal.pone.0157112

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* 26, 1045–1057. doi: 10.1007/s10278-013-9622-7

Ge, C., Gu, I. Y. H., Jakola, A. S., and Yang, J. (2020). Deep semi-supervised learning for brain tumor classification. *BMC Med. Imaging* 20:87. doi: 10.1186/s12880-020-00485-0

Ghasemi, M., Kelarestaghi, M., Eshghi, F., and Sharifi, A. (2020). T2-FDL: a robust sparse representation method using adaptive type-2 fuzzy dictionary learning for medical image classification. *Expert Syst. Appl.* 158:113500. doi: 10.1016/j.eswa.2020.113500

Ghassemi, N., Shoeibi, A., and Rouhani, M. (2020). Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. *Biomed. Signal Process. Control* 57:101678. doi: 10.1016/j.bspc.2019.101678

Gu, X., Zhang, C., and Ni, T. (2020). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 18, 53–61. doi: 10.1109/TCBB.2020.2973978

Hua, L., Gu, Y., Gu, X. Q., Xue, J., and Ni, T. G. (2021). A novel brain MRI image segmentation method using an improved multi-view fuzzy c-means clustering algorithm. *Front. Neurosci.* 15:662674. doi: 10.3389/fnins.2021.662674

Ismael, M. R., and Abdel-Qader, I. (2018). "Brain tumor classification via statistical features and back-propagation neural network," in *Proceedings of the 2018 IEEE International Conference on Electro/Information Technology (EIT)*, Rochester, MI, 252–257. doi: 10.1109/EIT.2018.8500308

Jiang, Y., Gu, X., Ji, D., Qian, P., Xue, J., Zhang, Y., et al. (2020). Smart diagnosis: a multiple-source transfer TSK fuzzy system for EEG seizure identification. *ACM Trans. Multimed. Comput. Commun. Appl.* 16:59. doi: 10.1145/3340240

Jiang, Y., Gu, X., Wu, D., Hang, W., Xue, J., Qiu, S., et al. (2021). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. *IEEE/ACM Trans. Comput.Biol. Bioinformatics* 18, 40–52. doi: 10.1109/TCBB.2019.2963873

Jiang, Z., Lin, Z., and Davis, L. S. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88

Kiranmayee, B. V., Rajinikanth, T. V., and Nagini, S. (2016). Enhancement of SVM based MRI brain image classification using pre-processing techniques. *Indian J. Sci. Technol.* 9, 1–7. doi: 10.17485/ijst/2016/v9i29/91042

Liu, Y. H., Muftah, M., Das, T., Bai, L., Robson, K., and Auer, D. (2012). Classification of MR tumor images based on Gabor wavelet analysis. *J. Med. Biol. Eng.* 32, 22–28. doi: 10.5405/jmbe.813

Long, M., Wang, J., Ding, G., Sun, J., and Yu, P. S. (2013). "Transfer feature learning with joint distribution adaptation," in *Proceedings of the International Conference on Computer Vision*, (New York, NY: IEEE), 2200–2207. doi: 10.1109/ICCV.2013.274

Mallick, P. K., Ryu, S. H., Satapathy, S. K., Mishra, S., Nguyen, G. N., and Tiwari, P. (2019). Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network. *IEEE Access* 15, 46278–46287. doi: 10.1109/ACCESS.2019.2902252

Mittal, M., Goyal, L. M., Kaur, S., Kaur, I., Verma, A., and Hermanth, D. J. (2019). Deep learning based enhanced tumor segmentation approach for MR brain images. *Appl. Soft Comput. J.* 78, 346–354. doi: 10.1016/j.asoc.2019.02.036

Mohankumar, S. (2016). Analysis of different wavelets for brain image classification using support vector machine. *Int. J. Adv. Signal Image Sci.* 30, 1–4. doi: 10.29284/ijasis.2.1.2016.1-4

Ni, T., Gu, X., and Jiang, Y. (2020). Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy. *J. Ambient Intell. Humaniz. Comput.* doi: 10.1007/s12652-020-02620-9 [Epub ahead of Print].

Ni, T., Gu, X., Wang, H., Zhang, Z., Chen, S., and Jin, C. (2018a). Discriminative deep transfer metric learning for cross-scenario person reidentification. *J. Electron. Imaging* 27:043026. doi: 10.1117/1.JEI.27.4.043026

Ni, T., Gu, X., Wang, J., Zheng, Y., and Wang, H. (2018b). Scalable transfer support vector machine with group probabilities. *Neurocomputing* 273, 570–582. doi: 10.1016/j.neucom.2017.08.049

Nikam, P. B., and Shinde, V. D. (2013). MRI brain image classification and detection using distance classifier method in image processing. *Int. J. Eng. Res. Technol.* 2, 1980–1985.

Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., and Ahuja, C. K. (2016). A package-SFERCB-"segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors". *Appl. Soft Comput.* 47, 151–167. doi: 10.1016/j.asoc.2016.05.020

Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., and Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data. *J. Comput. Sci.* 30, 174–182. doi: 10.1016/j.jocs.2018.12.003

Singhal, V., Aggarwal, H. K., Tariyal, S., and Majumdar, A. (2017). Discriminative robust deep dictionary learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 55, 5274–5283. doi: 10.1109/TGRS.2017.2704590

Song, J., Xie, X., Shi, G., and Dong, W. (2019). Multi-layer discriminative dictionary learning with locality constraint for image classification. *Pattern Recognit.* 91, 135–146. doi: 10.1016/j.patcog.2019.02.018

Sun, R., Wang, K., Guo, L., Yang, C., Chen, J., Ti, Y., et al. (2019). A potential field segmentation based method for tumor segmentation on multi-parametric MRI of glioma cancer patients. *BMC Med. Imaging* 19:48. doi: 10.1186/s12880-019-0348-y

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. *Cancer J. Clin.* 71, 1–41. doi: 10.3322/caac.21660

Tuia, D., and Camps-Valls, G. (2016). Kernel manifold alignment for domain adaptation. *PLoS One* 11:e0148655. doi: 10.1371/journal.pone.0148655

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 210–227. doi: 10.1109/TPAMI.2008.79

Zeng, K., Zheng, H., Cai, C., Yang, Y., Zhang, K., and Chen, Z. (2018). Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Comput. Biol. Med.* 99, 133–141. doi: 10.1016/j.compbiomed.2018.06.010

Zhang, L., Yang, M., and Feng, X. (2011). "Sparse representation or collaborative representa- tion: which helps face recognition?," in *Proceedings of the IEEE Conference on Computer Vision*, Vol. 2011, Barcelona, 471–478. doi: 10.1109/ICCV.2011.6126277

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fnins.2021.679847

Brain Tumor MR Image Classification Using Convolutional Dictionary Learning With Local Constraint

*Xiaoqing Gu[1], Zongxuan Shen[1], Jing Xue[2], Yiqing Fan[3] and Tongguang Ni[1]**

[1] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China, [2] Department of Nephrology, Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China, [3] Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

Brain tumor image classification is an important part of medical image processing. It assists doctors to make accurate diagnosis and treatment plans. Magnetic resonance (MR) imaging is one of the main imaging tools to study brain tissue. In this article, we propose a brain tumor MR image classification method using convolutional dictionary learning with local constraint (CDLLC). Our method integrates the multi-layer dictionary learning into a convolutional neural network (CNN) structure to explore the discriminative information. Encoding a vector on a dictionary can be considered as multiple projections into new spaces, and the obtained coding vector is sparse. Meanwhile, in order to preserve the geometric structure of data and utilize the supervised information, we construct the local constraint of atoms through a supervised *k*-nearest neighbor graph, so that the discrimination of the obtained dictionary is strong. To solve the proposed problem, an efficient iterative optimization scheme is designed. In the experiment, two clinically relevant multi-class classification tasks on the Cheng and REMBRANDT datasets are designed. The evaluation results demonstrate that our method is effective for brain tumor MR image classification, and it could outperform other comparisons.

Keywords: brain tumor image classification, magnetic resonance imaging, dictionary learning, local constraint, convolutional neural network

OPEN ACCESS

Edited by:
Mohammad Khosravi, Persian Gulf University, Iran

Reviewed by:
Min Shi, Fuzhou University of International Studies and Trade, China
Yanhui Zhang, Hebei University of Chinese Medicine, China

****Correspondence:***
Tongguang Ni
ntg@cczu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *12 March 2021*
Accepted: *09 April 2021*
Published: *28 May 2021*

Citation:
Gu X, Shen Z, Xue J, Fan Y and Ni T (2021) Brain Tumor MR Image Classification Using Convolutional Dictionary Learning With Local Constraint. Front. Neurosci. 15:679847. doi: 10.3389/fnins.2021.679847

INTRODUCTION

Brain tumors are abnormal cell aggregations that grow inside the brain tissues. Brain tumors can be divided into benign tumors and malignant tumors. Brain benign tumors can be cured by surgery, while malignant brain tumors are one of the most deadly types of cancer and can lead directly to death (Yang et al., 2018; Sun et al., 2019; Ge et al., 2020). Brain tumors can also be divided into primary tumors formed in the brain or derived from the brain nerves and metastatic brain tumors metastasized from other parts of the body to the brain. The most common primary brain tumors in adults are primary central nervous system lymphoma and gliomas, of which gliomas originate from the periglial tissue and account for more than 80% of malignant brain tumors. Different symptoms appear with different lesion areas, such as headache, vomiting, visual decline, epilepsy, and confusion. A more detailed classification divides brain tumors into four grades, the higher the grade, the more malignant the brain tumor is. According to the Global cancer statistics 2020 (Sung et al., 2021), there are about 308,000 new cases of brain cancers in 2020, accounting for about 1.6%

of all new cases of cancers, and about 251,000 deaths from brain cancers, accounting for about 2.5% of all cancer deaths.

Early detection is important for effective treatment of brain tumors (Gumaei et al., 2019). With the development of medical imaging, imaging techniques play an important role in brain tumor diagnosis and treatment evaluation and can provide doctors with a clear human brain structure. These imaging techniques can provide information on the shape, size, and location of brain tumors, assisting doctors to make an accurate diagnosis and develop a treatment plan. Magnetic resonance (MR) imaging is one of the most commonly used scanning methods in neurology. MR imaging uses radiofrequency signals to excite the target tissue under the influence of a very strong magnetic field to produce an image of its interior. It has the advantages of high soft tissue contrast and zero ionizing radiation exposure. Therefore, MR imaging is more suitable for the detection of brain lesions (Zeng et al., 2018; Mittal et al., 2019; Bunevicius et al., 2020).

In recent years, artificial intelligence has attracted more and more attention due to its achievements in the field of intelligent medicine. The classification and segmentation of MR images using artificial intelligence methods has become a hot topic in the research of medical image processing (Mohan and Subashini, 2018; Anaraki et al., 2019). The application of brain tumor classification falls into two main types: classification of brain images into normal and abnormal, that is, whether the brain image contains a tumor or not; and classification within abnormal brain images, that is, differentiation between different classes of brain tumors. Classifying brain tumors into different pathological classes is more challenging than a binary classification. The challenge lies in brain tumors being permeable, their appearance is highly heterogeneous, their location is random, and the number of voxels in each subregion varies widely (Chahal et al., 2020).

Brain tumor classification includes two procedures: feature extraction and classification. In some previous studies, traditional manual extraction of features was widely used, such as intensity and texture features of brain tumor images. However, traditional feature extraction methods require the professional knowledge and experience in specific fields. Manual feature extraction will also reduce the efficiency of the system. Deep learning techniques overcome this disadvantage (Sajjad et al., 2019). Feature extraction methods based on deep learning have demonstrated successful results in real-world medical image processing applications (Deepak and Ameer, 2019). Among various classification methods, dictionary learning (DL) is a powerful tool in image processing and machine vision, making sparse coding tasks efficient and robust (Al-Shaikhli et al., 2016; Ni et al., 2020). The sparse coding can approximate high-dimensional image features into a linear combination of a few atoms from the learned dictionary (Li et al., 2018; Gu et al., 2020b). Ghasemi et al. (2020, 2021) developed fuzzy dictionaries to deal with the uncertainty in brain tumor image classification. The classic fuzzy inference is embedded into the dictionary learning process and fuzzy membership functions are used to model uncertainty and improve sparse representation. Wu et al. (2017) developed a parse representation method to exact important features and key feature index across different class images. Then, the learned feature weights and classification dictionary are used in a radiomics system for the diagnosis of brain tumors. Al-Shaikhli et al. (2014) developed a coupled dictionary learning method, which designs one dictionary of brain tumor image patches and one dictionary of image labels. The label dictionary is used to present the foreground and background multiple labels. Then, Al-Shaikhli et al. (2016) extended this work by using the information of brain topology and texture to develop a multi-class brain tumor classification method. Chen et al. (2017) proposed a kernel sparse representation method for multi-label brain tumor segmentation. This method consists of three main parts as principal component analysis—split for dictionary learning initialization, second for kernel sparse representation processing of kernel dictionary learning and kernel sparse coding, and third for making brain image segmentation using graph-cut method. Adebileje et al. (2017) applied two dictionary learning methods for classifying proton MR spectroscopy of brain gliomas tumor, i.e., one is discriminate sub-dictionary learning method, and the other is projective dictionary pair learning. These two methods were tested on many H-MRS patients signal samples selected in an Iran hospital and evaluated to be noise insensitive. Tong et al. (2019) proposed a kernel dictionary learning method that segmented MR brain tumor images after the noise removal and contrast enhancement and then extracted the nonlinear features by the learned kernel dictionary for healthy and pathologically tissues. Finally, the segmentation is done by kernel-clustering method. Vu et al. (2015) proposed a feature-oriented dictionary learning method. This method incorporated feature extraction discriminative into dictionary learning. In addition, it built discriminative class-specific dictionaries that emphasized the small intra-class differences and large inter-class differences. Finally, this method had been evaluated using brain cancer dataset Cancer Genome Atlas.

Our goal in this study is to build an automatic and effective brain tumor MR image classification method to assist physicians in decision-making. In order to capture the better discriminative feature representations of brain tumor MR images, we propose convolutional dictionary learning with local constraint (CDLLC) method to seek sparse feature representation and dictionary simultaneously by using a convolutional neural network (CNN) framework. In addition, we employ the locality constraint term on codes in the last layer. The locality constraint term is used to enforce the manifold structure of the codes to preserve the locality information. Various CNN structures can be used in CDLLC; in this study, we use AlexNet (Krizhevsky et al., 2012) and softmax classifier loss in the last layer. The framework of the proposed method is shown in **Figure 1**. The advantages of the proposed CDLLC method are as follows: (1) CDLLC learns a multi-layer convolutional dictionary for feature representation and encoding in the nonlinear space, so that the nonlinear latent information of data is employed. (2) Encoding a vector on a multi-layer dictionary can be considered as multiple projections into new spaces. The projection is nonlinear and the obtained coding vector is sparse. Simultaneously, the resulting coding vectors of different classes can give the discriminative

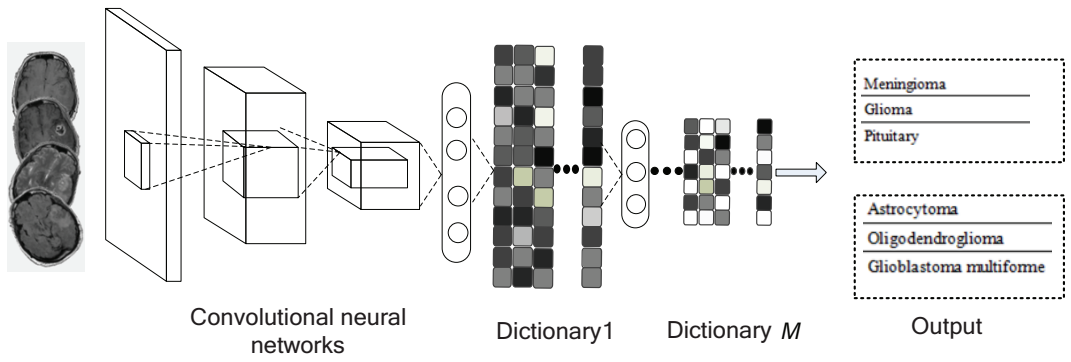


FIGURE 1 | Framework of the proposed method.

approximation and delete the redundant information. For example, the coding vectors in single layer dictionary learning may be nonlinear separable, and they will be transformed into linear separable in our method. (3) By considering the supervised information and graph Laplacian regularization term, the learned coding vectors are more discriminative. Simultaneously, graph Laplacian regularization preserves the locality structure information of the learned dictionary in the last layer. (4) The proposed CDLLC method is conducted on two public brain tumor datasets. The performance and usefulness of CDLLC are validated in terms of accuracy, recall, precision, F1-score, and balance loss.

The rest of the article is organized as follows: the related work is introduced in section "Backgrounds." The proposed method is given in section "Convolutional Dictionary Learning With Local Constraint," and experiments are reported in section "Experiments." Finally, a conclusion is summarized in section "Conclusion."

BACKGROUNDS

Dataset

The brain tumor datasets used in this article are provided by Cheng et al. (2015, 2016) and the Repository of Molecular Brain Neoplasia Data (REMBRANDT) (Clark et al., 2013). The images provided by Cheng are 3064 T1-weighted contrast-enhanced images, containing 708 meningiomas, 1426 gliomas, and 930 pituitary tumors. All images are digitized at a resolution of 512 × 512 pixels. The REMBRANDT dataset contains 110,020 pre-surgical MR multi-sequence images from 130 brain tumor patients. The dataset contains astrocytoma (AST), oligodendroglioma (OLI), glioblastoma multiforme (GBM), and other unidentified tumor types. All images are digitized at a resolution of 256 × 256 pixels. Each image in the Cheng and REMBRANDT datasets is labeled with one type of brain tumor. The example samples of the Cheng and REMBRANDT datasets are shown in **Figures 2**, **3**, respectively. The challenge of these two datasets lies in some factors, such as high variability in shape, size, and the similar presentation of different pathological types.

Dictionary Learning

Let $X=[x_1,x_2,\ldots,x_N]\in R^{d\times N}$ be the labeled training images and $A \in R^{K\times N}$ be the sparse coding vector matrix. Denote the dictionary to be learned by $D\in R^{d\times K}$. Consider the linear representation, $DA\approx X$. The dictionary **D** can be learned as

$$\begin{aligned} &\min_{D,A} \ \| \text{X-DA} \|_F^2 + \lambda\Theta(A), \\ &s.t. \ \| d_i \|_2^2 \leq 1, \ \forall i \end{aligned} \tag{1}$$

where the first term is reconstruction error and $\Theta(A)$ represents the constraints of coding vector matrix **A**, such as ℓ_0, ℓ_1, ℓ_2 and Frobenius norm. The parameter λ is a positive scalar, which controls the sparsity. $\|d_i\|_2^2\leq 1$ is used to control the complexity of model. Its purpose is to prevent dictionary **D** from being arbitrarily large, as it will result in very small values for coding matrix **A**. Parameters **D** and **A** are often optimized by alternating iterations until convergence.

Equation 1 is an unsupervised learning framework. It can obtain good performance in reconstruction tasks and can also be used in some classification tasks, as it is good at mining potential patterns in the data. In order to make better use of supervised information in classification tasks, different kinds of loss functions are considered in dictionary learning. The supervised dictionary learning can be presented as

$$\begin{aligned} &\min_{\theta,D} \ \textstyle\sum_{x\in X} L(l_x, a_x, D, \theta), \\ &a_x = \arg\min \ \| l_x - Da \|_2^2 + \lambda\Theta(a), \end{aligned} \tag{2}$$

where l_x is the class label of training sample **x**. Determining the suitable classification loss function L and its parameter θ are critical to classification tasks. The joint learning of parameters **D** and θ allows **D** to have discriminative capability with minimal classification cost. In order to obtain the optimal solution of Eq. 2, gradient descent method, back propagation method, or orthogonal matching pursuit (OMP) (Wang et al., 2012; Peng et al., 2020) can be used.

Convolution Dictionary Learning

Convolution dictionary learning has been proposed in recent years (Papyan et al., 2018; Sulam et al., 2018; Song et al., 2019).

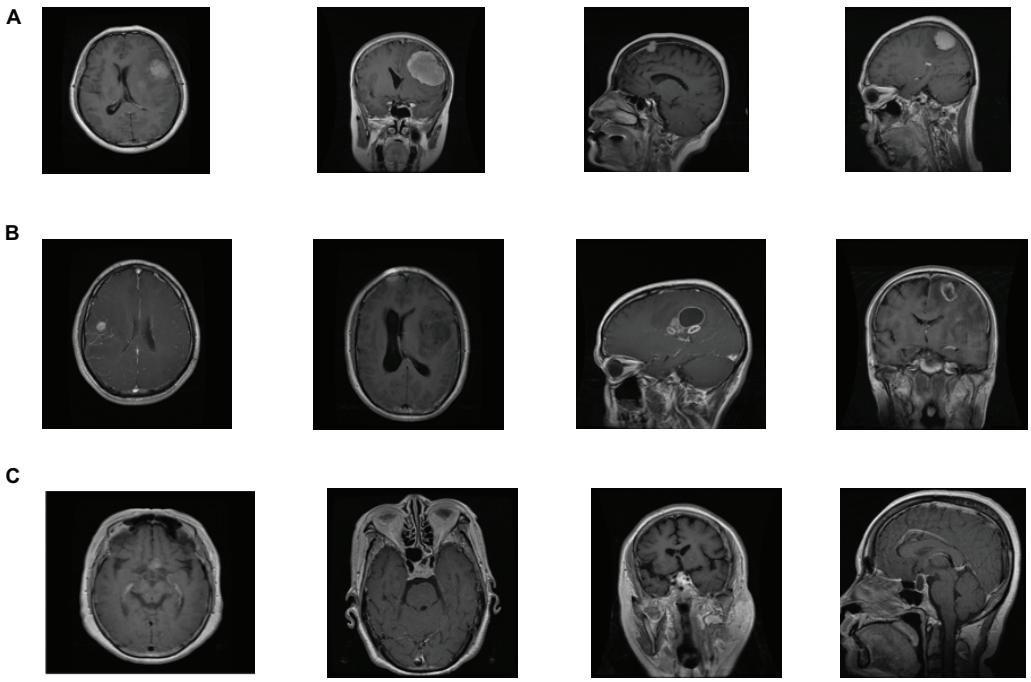


FIGURE 2 | Example samples of Cheng dataset: **(A)** meningioma, **(B)** glioma, **(C)** and pituitary tumor.

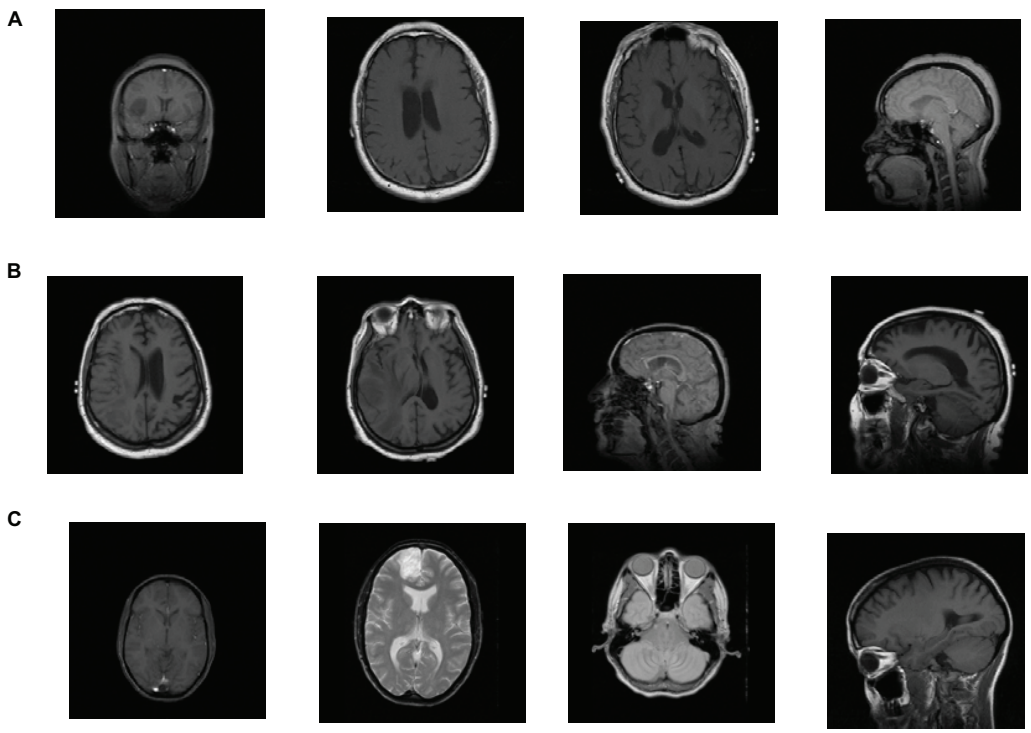


FIGURE 3 | Example samples of REMBRANDT dataset: **(A)** AST, **(B)** GBM, and **(C)** OLI.

The model is an extension of traditional dictionary learning. Its aim is to capture the deep structure of data and increase the discriminability of features. Convolution dictionary learning follows the architecture of CNN and is used in a hierarchical way. The convolution of the filter in CNN corresponds to the sparse coding step in multi-layer convolution dictionary. Let

$\{D_m\}_{m=1}^{M}$ be M-layer convolution dictionary, where $D_m \in R^{d\times K_m}$ is the dictionary in the m-th layer dictionary and K_m is the size of D_m. The convolutional representation of $\mathbf{X}$ can be presented as $X \approx D_1D_2...D_MA_M$. In detail, with the decomposition constraint $D_{m-1} = D_mA_m$, this process can be described as

$$\begin{aligned} X &\approx D_1A_1, \\ X &\approx D_1D_2A_2, \\ X &\approx D_1D_2D_3A_3, \\ &\vdots \\ X &\approx D_1D_2...D_MA_M. \end{aligned} \tag{3}$$

where $A_m \in R^{K_m\times N}$ is the coding vector matrix in the m-th layer.

CONVOLUTIONAL DICTIONARY LEARNING WITH LOCAL CONSTRAINT

Objective Function

Consider multi-layer dictionary architecture with M layers, the coding A_m in the m-th layer can be written as

$$A_m \approx \phi(D_{m+1}A_{m+1}), \tag{4}$$

where ϕ is a nonlinear function, such as rectified linear unit (ReLU), Sigmoid, and TanHyperbolic activation function. In this case, coding using a dictionary is used as a projection into another feature space, and the coding vector is a new input for the next layer. In order to preserve the essential structure information of data, it is essential to reconstruct the original sample in the last layer. We consider the following equation:

$$\begin{aligned} J_1(D_1, D_2, ..., D_M, A_M) = &\parallel \text{X-D}_1\phi(D_2\phi(...\phi(D_MA_M))) \parallel_F^2 \\ &+\lambda\Theta(A_M), \end{aligned} \tag{5}$$

where λ is the regularization parameter.

Following Cai et al. (2013) and Ding and Fu (2018), we use the rank operator for $\Theta(A_M)$. In this case, A_M can be approximated as $A_M \approx SH$, and the $\Theta(A_M)$ term can be written as

$$\parallel A_M - SH \parallel_F^2, \tag{6}$$

where $S \in R^{K_M\times C}$ and $H \in R^{C\times N}$. C is the class number of data samples.

To improve the classification performance, local information takes an important part in dictionary learning (Peng et al., 2020). Since the atoms of dictionary is more robust and stable than original samples, we use a graph Laplacian regularization term of atoms to trace the manifold structure of data. We build a supervised k-nearest neighbor graph $\mathbf{W}$ of atoms on dictionary D_M in the last layer. The element $w_{i,j}$ in graph $\mathbf{W}$ is denoted as follows:

$$w_{i,j} = \begin{cases} \exp\left(-\dfrac{\parallel d_{M,i} - d_{M,j} \parallel^2}{\sigma}\right), & \text{if} \quad d_{M,i} \in KNN(d_{M,j}) \\ & \text{and} \quad l_i = l_j \\ 0, & \text{else} \end{cases}, \tag{7}$$

where σ is an adjustable parameter. Different from the unsupervised graph Laplacian (Peng et al., 2020), we embed the supervised information in graph $\mathbf{W}$, such that similar codes are enforced and more discriminative information can be learned.

Then, we construct the graph Laplacian regularization term in the last layer as follows:

$$J_2(A_M) = \sum_{i=1}^{K_M}\sum_{j=1}^{K_M} w_{i,j} \parallel a_{M,i} - a_{M,j} \parallel_2^2 = Tr(A_M^T L A_M), \tag{8}$$

where the Laplacian matrix $L = diag(w_1, w_2, ..., w_{K_M}) - W$, and $w_i = \sum_{j=1}^{K_M} w_{i,j}$.

To further learn a discriminative dictionary by exploiting supervised information, we use the softmax classifier loss in the last layer, which is commonly used in CNN structure

$$J_3(A_M, \theta) = -\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{C} p_{c,i} log \frac{e^{\theta_c^T a_{M,i}}}{\sum_{u=1}^{C} e^{\theta_u^T a_{M,i}}}, \tag{9}$$

where $p_{c,i}$ is the label probability of $a_{M,i}$ and is assigned to class c. $\theta = [\theta_1, \theta_2, ..., \theta_C]$ is the adjustable parameter in softmax classifier.

For brain tumor image classification, we combine the three functions J_1, J_2, and J_3 together, jointly optimizing convolutional dictionary learning and classifier. Therefore, we have the objective function of the CDLLC model

$$\begin{aligned} &\arg\min J_1(D_1, D_2, ..., D_M, A_M) + J_2(A_M) + J_3(A_M, \theta), \\ &\quad s.t. \ \parallel d_i \parallel_2^2 \leq 1, \ \forall i \end{aligned} \tag{10}$$

TABLE 1 | Confusion matrix of CDLLC on the Cheng dataset.

	Meningioma	Glioma	Pituitary
Meningioma	0.8875	0.0782	0.0343
Glioma	0.0444	0.9487	0.0069
Pituitary	0.0093	0.0070	0.9837

TABLE 2 | Classification performance of CDLLC on the Cheng dataset.

		Accuracy	Recall	Precision	F1-score	Balance loss
Fold-1	Training	96.94	94.97	95.17	94.95	96.56
	Test	96.36	94.59	94.78	94.70	96.03
Fold-2	Training	96.76	95.15	94.78	95.25	96.44
	Test	96.29	94.71	94.42	94.63	96.08
Fold-3	Training	96.77	95.02	95.21	95.30	96.41
	Test	96.32	94.62	94.51	94.65	96.11
Fold-4	Training	96.83	94.89	94.99	95.10	96.47
	Test	96.35	94.67	94.60	94.75	96.06
Fold-5	Training	97.12	95.09	95.28	95.17	96.96
	Test	96.39	94.64	94.61	94.70	96.22

To show all terms clearly, we expand the expression as follows:

$$
\begin{aligned}
&[D_1, D_2, ..., D_M, A_M, \theta] \\
&= \arg\min \| \mathrm{X}\text{-}\mathrm{D}_1\phi(D_2\phi(...\phi(D_M A_M))) \|_F^2 + \lambda_1 \| A_M - SH \|_F^2 \\
&\quad + \lambda_2 Tr(A_M^T L A_M) - \frac{\lambda_3}{M} \sum_{i=1}^{M} \sum_{c=1}^{C} p_{c,i} \log \frac{e^{\theta_c^T a_{M,i}}}{\sum_{u=1}^{C} e^{\theta_u^T a_{M,i}}}, \\
&\quad s.t. \ \| d_i \|_2^2 \leq 1, \ \forall i
\end{aligned} \tag{11}
$$

This jointly optimizing convolutional dictionary learning and classifier has benefits. During the procedure of optimizing, the convolutional dictionary learning gradually enhances the classification performance of the classifier; meanwhile, the learned classifier also improves the numerical stability and discriminative ability of dictionary coding.

Optimization of CDLLC

The optimization of Eq. 11 is not convex, and we solve dictionaries $\{D_1, D_2, ..., D_M\}$, coding matrix $\mathbf{A}$ and classifier parameter θ by an alternative optimization approach. In each step of iteration, we compute a certain parameter and fix the other parameters.

When fixing $\mathbf{A}$ and θ, we update dictionaries $\{D_1, D_2, ..., D_M\}$. We use the chain rule to compute $D_m (1 \leq m \leq M)$ in each layer

$$
\begin{aligned}
\frac{\partial J}{\partial D_m} &= \frac{\partial J}{\partial (D_m A_m)} A_m^T \\
&= \left[\frac{\partial J}{\partial \phi(D_m A_m)} \odot \phi'(D_m A_m) \right] A_m^T \\
&= \left[\frac{\partial J}{\partial A_{m-1}} \odot \phi'(D_m A_m) \right] A_m^T,
\end{aligned} \tag{12}
$$

where $\odot$ denotes the element-wise multiplication. Specifically, we can obtain $\frac{\partial J}{\partial D_1} = 2(D_1 A_1 - X) A_1^T$. After dictionary D_M is obtained, we use Eq. 7 to construct the graph Laplacian regularization.

When fixing $\{D_1, D_2, ..., D_M\}$ and θ, we update coding matrix $A_m (1 < m \leq M-1)$ in each layer as

$$
\begin{aligned}
\frac{\partial J}{\partial A_m} &= (D_m)^T \frac{\partial J}{\partial (D_m A_m)} \\
&= (D_m)^T \left[\frac{\partial J}{\partial \phi(D_m A_m)} \odot \phi'(D_m A_m) \right] \\
&= (D_m)^T \left[\frac{\partial J}{\partial A_{m-1}} \odot \phi'(D_m A_m) \right].
\end{aligned} \tag{13}
$$

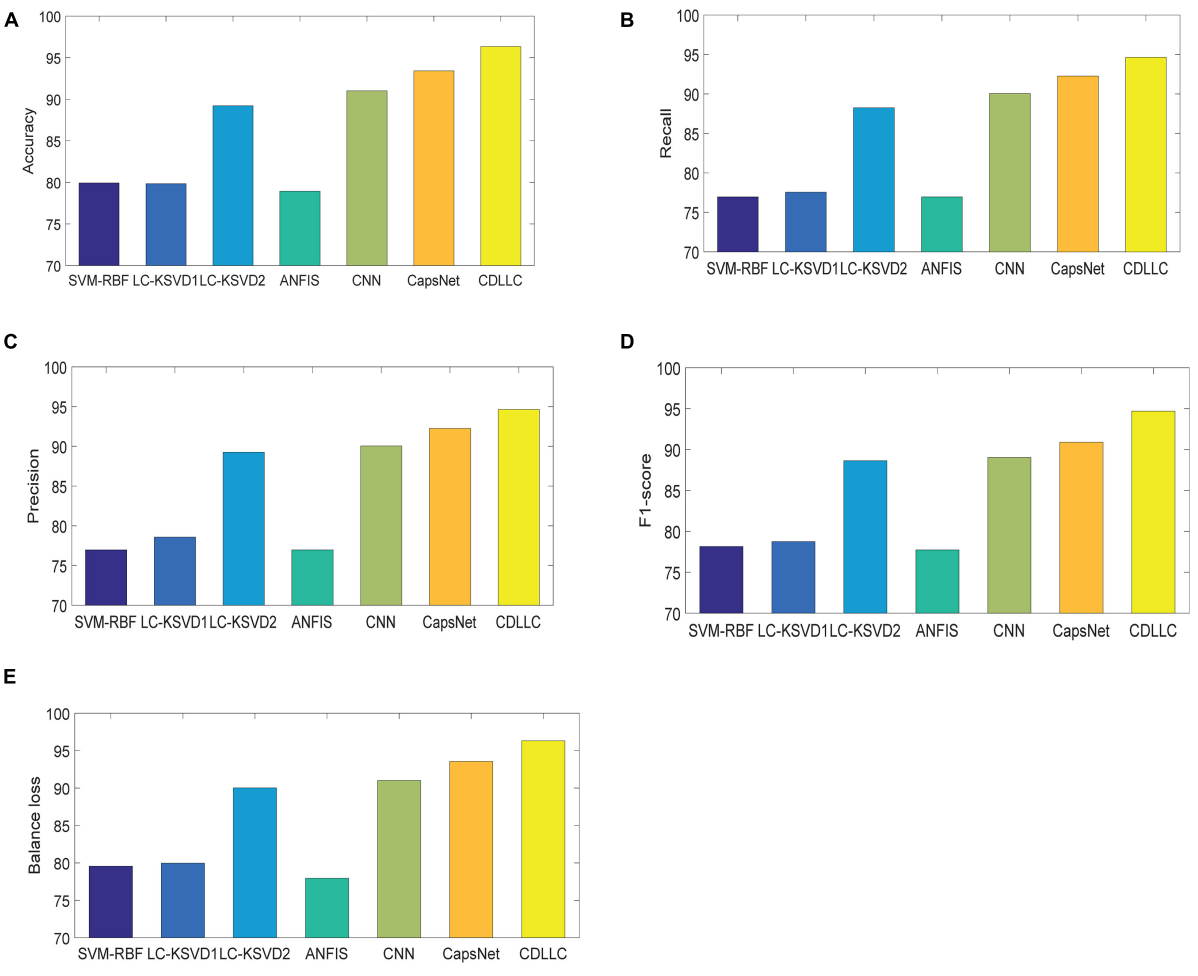


FIGURE 4 | Performance of CDLLC on the Cheng dataset: **(A)** accuracy, **(B)** recall, **(C)** precision, **(D)** F1-score, and **(E)** balance loss.

Specifically, we can obtain A_1 as

$$\frac{\partial J}{\partial A_1} = 2D_1^T(D_1A_1 - X). \quad (14)$$

The partial derivatives of the J regarding the A_M is given as

$$\frac{\partial J}{\partial A_M} = 2(D_M)^T\left[\frac{\partial J}{\partial A_{M-1}} \odot \phi'(D_MA_M)\right] + 2\lambda_1(A_M - SH) + 2\lambda_2 L_M A_M - \frac{\lambda_3}{M}\sum_{c=1}^{C}\left(\theta_c - \frac{\sum_{u=1}^{C}\theta_u e^{\theta_u^T A_{M,i}}}{\sum_{u=1}^{C} e^{\theta_u^T A_{M,i}}}\right). \quad (15)$$

Considering Eq. 6, we need to update matrixes **S** and **H** in turn. The optimal **S** can be computed as

$$\frac{\partial J}{\partial S} = A_M H^T - SHH^T. \quad (16)$$

Then, we can update **S** as

$$S = A_M H^T (HH^T)^\dagger, \quad (17)$$

where † is the Moore–Penrose pseudoinverse.

Similarly, the optimal **H** can be computed as

$$\frac{\partial J}{\partial H} = S^T SH - S^T A_M. \quad (18)$$

Then, we can update **S** as

$$H = (SS^T)^\dagger S^T A_M. \quad (19)$$

In this study, we use the softmax classifier for brain tumor image classification. The optimal classifier parameter θ can be computed as

$$\frac{\partial J}{\partial \theta_c} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{C} l_{c,i}\left(1 - \frac{e^{\theta_c^T a_{M,i}}}{\sum_{u=1}^{C} e^{\theta_u^T a_{M,i}}}\right)\alpha_{M,i}^T. \quad (20)$$

TABLE 3 | Confusion matrix of CDLLC on the REMBRANDT dataset.

	AST	OLI	GBM
AST	0.9686	0.0100	0.0214
OLI	0.0621	0.9127	0.0252
GBM	0.0583	0.0108	0.9309

TABLE 4 | Classification performance of CDLLC on the REMBRANDT dataset.

		Accuracy	Recall	Precision	F1-score	Balance loss
Fold-1	Training	97.74	93.89	95.42	94.25	95.86
	Test	97.55	93.67	95.28	94.10	95.39
Fold-2	Training	97.80	93.92	95.53	94.34	95.91
	Test	97.74	93.90	95.38	94.21	95.39
Fold-3	Training	97.87	93.96	95.52	94.35	95.82
	Test	97.74	93.89	95.39	94.22	95.56
Fold-4	Training	97.85	93.97	95.58	94.47	95.99
	Test	97.80	93.92	95.52	94.34	95.71
Fold-5	Training	97.74	93.90	95.48	94.26	95.79
	Test	97.37	93.66	95.12	94.05	95.42

Testing Brain Tumor MR Images

Let **x** be the feature descriptor of a test image, based on the learned optimal **S**, **H**, θ, and $\{D_1,D_2,...,D_M\}$, we can compute the dictionary encoding of **x** by the following formulation:

$$\min_{a_M} \| \text{x-D}_1\phi(D_2\phi(...\phi(D_Ma_M))) \|_F^2 + \lambda_1 \| a_M - SH \|_F^2 . \quad (21)$$

Finally, the softmax classifier is used for classification, and the label probability of **x** assigned to class c can be computed as

$$p_c = \frac{e^{\theta_u^T a_M}}{\sum_{u=1}^{C} e^{\theta_u^T a_M}}. \quad (22)$$

EXPERIMENTS

Experimental Settings

In our experiment, we randomly select 1000 images in the REMBRANDT dataset and the whole Cheng dataset for simulation. Brain MR images are resized to 227 × 227 sizes to use for AlexNet. All convolution layers in CDLLC employ filters of size 3 × 3. Stochastic gradient descent (SGD) is used as an optimizer and TensorFlow is implemented in model training. ReLU is used as the activation function. The initialization of dictionary is performed from $\mathbf{D}_1$ and $\mathbf{A}_1$. We run K-SVD algorithm on each class of data and integrate subclass dictionary into the dictionary $\mathbf{D}_1$. Then, we obtain the $\mathbf{D}_m$ and $\mathbf{A}_m$ based on the learned dictionary and coding from the previous layer. The parameters λ_1, λ_2, and λ_3 in CDLLC are selected in the search grid {0.001, 0.01, 0.1, 1, 10}. We use the fivefold cross-validation to complete with valid and comparable results. The 70% of the training fold are used for training and 30% are used as validation. We perform experiments for 10 times and then record their average values. We summarize the performance of all comparative methods in terms of accuracy, F1-score, precision, recall, and balance loss (Gu et al., 2018, 2020a).

There are many categories of methods used for brain tumor MR image classification. We compare several traditional machine learning and deep learning methods in the experiments. The traditional machine learning methods include the support vector machine with RBF kernel (SVM-RBF) (Nandpuru et al., 2014), the dictionary learning method label consistent K-SVD (called LC-KSVD1 in the experiment) (Jiang et al., 2013), and the neural network fuzzy inference system (ANFIS) (Thirumurugan and Shanthakumar, 2016). The deep learning classification methods include CNN (Krizhevsky et al., 2012) and CapsNet (Afshar et al., 2018). For these traditional classification methods, feature selection is a key step, where first- and second-order statistical texture features are widely used in brain tumor image classification. In our experiment, we follow (Fujima et al., 2019; Zhang et al., 2019) and use the statistical texture features: mean, variance, standard deviation, skewness, kurtosis, contrast, energy, entropy, correlation, and homogeneity. In addition, to compare with LC-KSVD using deep features, we exploit features from CNN architecture AlexNet and apply these features into LC-KSVD (called LC-KSVD2 in the experiment). We conduct the experiments on a computer with Intel Xeon Processor E5-2620

v4 and 64 GB RAM. All methods are implemented on Python 2.7, using Keras library and Tensor Flow.

Experimental Results on Cheng Dataset

In this subsection, we observe the classification performance of CDLLC on the Cheng dataset. The images of the Cheng dataset used in our experiment contain three types of brain tumors: meningioma, glioma, and pituitary. First, we show the confusion matrix for three classes obtained by CDLLC in **Table 1**. The confusion matrix provides the valuable information about the predicted labels. From **Table 1**, we can see that the pituitary tumor is classified with the highest accuracy, glioma tumors are classified with the second accuracy, and meningioma tumors are classified with the lowest accuracy. Generally, the overall classification accuracy is satisfactory.

Second, we use accuracy, F1-score, precision, recall, and balance loss as the evaluation indexes. The performance of these five indexes of CDLLC in five folds is shown in **Table 2** in detail. **Table 2** shows that CDLLC obtains high average values and small standard deviation on accuracy, F1-score, precision, recall, and balance loss. Then, we compare our method with RBF-SVM, LC-KSVD1, LC-KSVD2, ANFIS, CNN, and CapsNet. The average experimental results on the Cheng dataset are shown in **Figure 4**. It can be observed from these results that (1) CDLLC achieves the best results in comparison with other methods. This indicates that more discriminative information can be exploited by the proposed method. It suggests that the multi-layer dictionary learning, which addresses both feature representation and encoding in the nonlinear space, can exploit discriminative features from deep learning structure. In addition, graph Laplacian regularization can preserve the locality structure information of sparse codes, which can largely improve the model discriminative ability. (2) Among all methods, deep leaning methods gain better classification performance than traditional machine learning (SVM-RBF, LC-KSVD1, and ANFIS) with statistical texture features. Using the deep features, the classification performance of LC-KSVD2 is obviously improved than LC-KSVD1. It indicates that deep features are more adapted to brain tumor image classification.

Experimental Results on REMBRANDT Dataset

In this subsection, we observe the classification performance of CDLLC on the REMBRANDT dataset. The images of the REMBRANDT dataset used in our experiment contain three types of brain tumors: AST, OLI, and GBM.

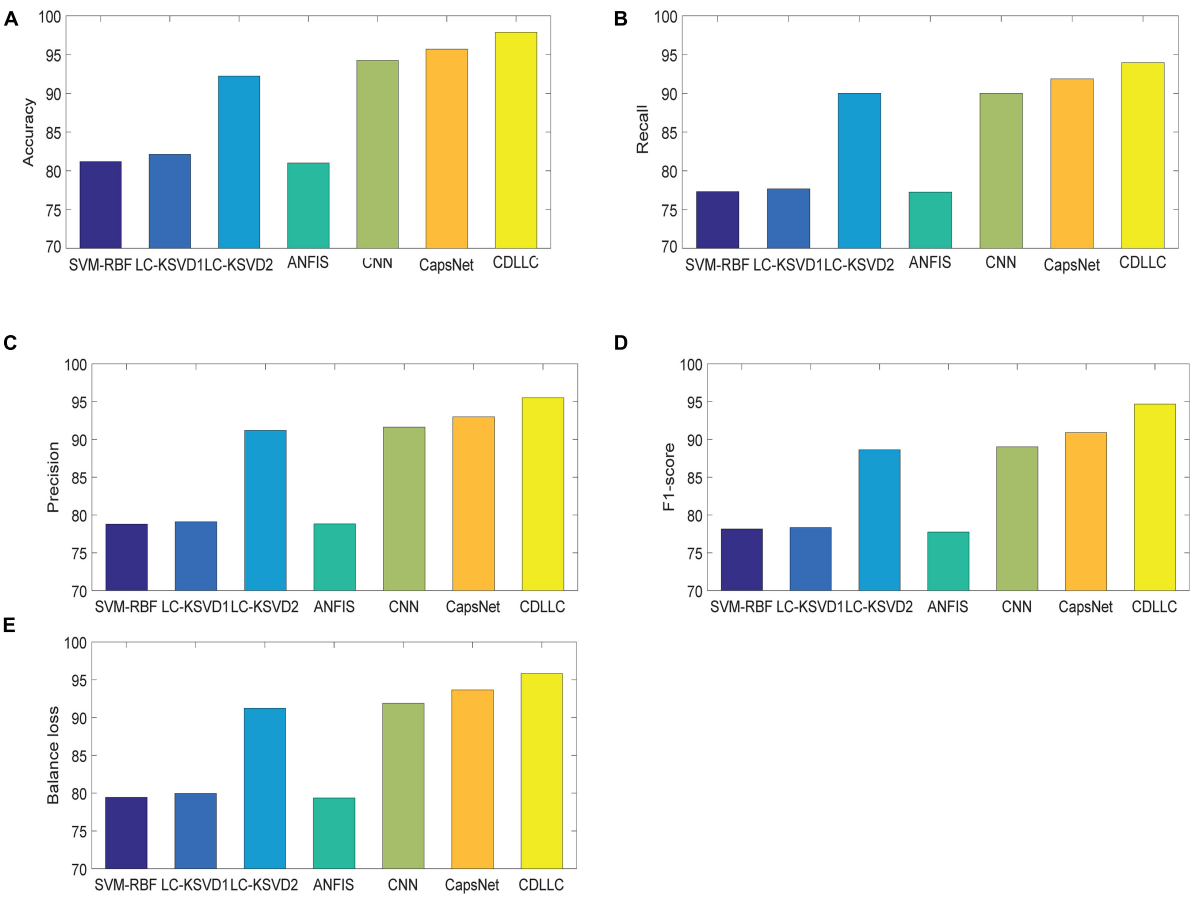


FIGURE 5 | Performance of CDLLC on the REMBRANDT dataset: **(A)** accuracy, **(B)** recall, **(C)** precision, **(D)** F1-score, and **(E)** balance loss.

We first show the confusion matrix for three classes obtained by CDLLC in **Table 3**. From **Table 3**, we can see that the classification performance of CDLLC for three types of brain tumor images is comparable. The classification rates of AST, OLI, and GBM are 0.9686, 0.9127, and 0.9309, respectively. Second, we summarize the performance of CDLLC in terms of accuracy, F1-score, precision, recall, and balance loss in **Table 4**. From the fivefold results in **Table 4**, we can see that CDLLC gains the satisfactory results on the REMBRANDT dataset. Our multi-layer dictionary structure is not only convolutional but also sparse, and all parameters are updated within joint optimization learning. Then, we compare our method with RBF-SVM, LC-KSVD, ANFIS, CNN, and CapsNet. The average experimental results on the REMBRANDT dataset are shown in **Figure 5**. Similar to the results in **Figure 4**, CDLLC gains the best performance in five evaluation indexes. LC-KSVD1 is as a baseline method of our method. Whether using statistical texture features or deep features, the performance of LC-KSVD1 is much lower than that of the proposed CDLLC. The reason is that LC-KSVD1 learns the dictionary in the original space and such dictionary cannot well exploit the discriminative information.

Parameter Analysis

In this section, we analyze the parameter sensitivity of CDLLC on the Cheng and REMBRANDT datasets. First, we discuss the parameters λ_1, λ_2, and λ_3 in CDLLC. These three parameters are selected in the search grid {0.001, 0.01,..., 10}. We set $\lambda_1=\lambda_2$ and visualize the change of classification accuracy of CDLLC with different values of λ_2 and λ_3. Similarly, we set $\lambda_2=\lambda_3$ ($\lambda_1=\lambda_3$) and visualize the change of classification accuracy of CDLLC with different values of λ_1 and λ_3 (λ_1 and λ_2). The results of classification accuracy are shown in **Figures 6**, **7**. We can see that different values of the parameters λ_1, λ_2, and λ_3 have a significant impact on the classification accuracy of CDLLC. It indicates that the grid search strategy is appropriate for λ_1, λ_2, and λ_3.

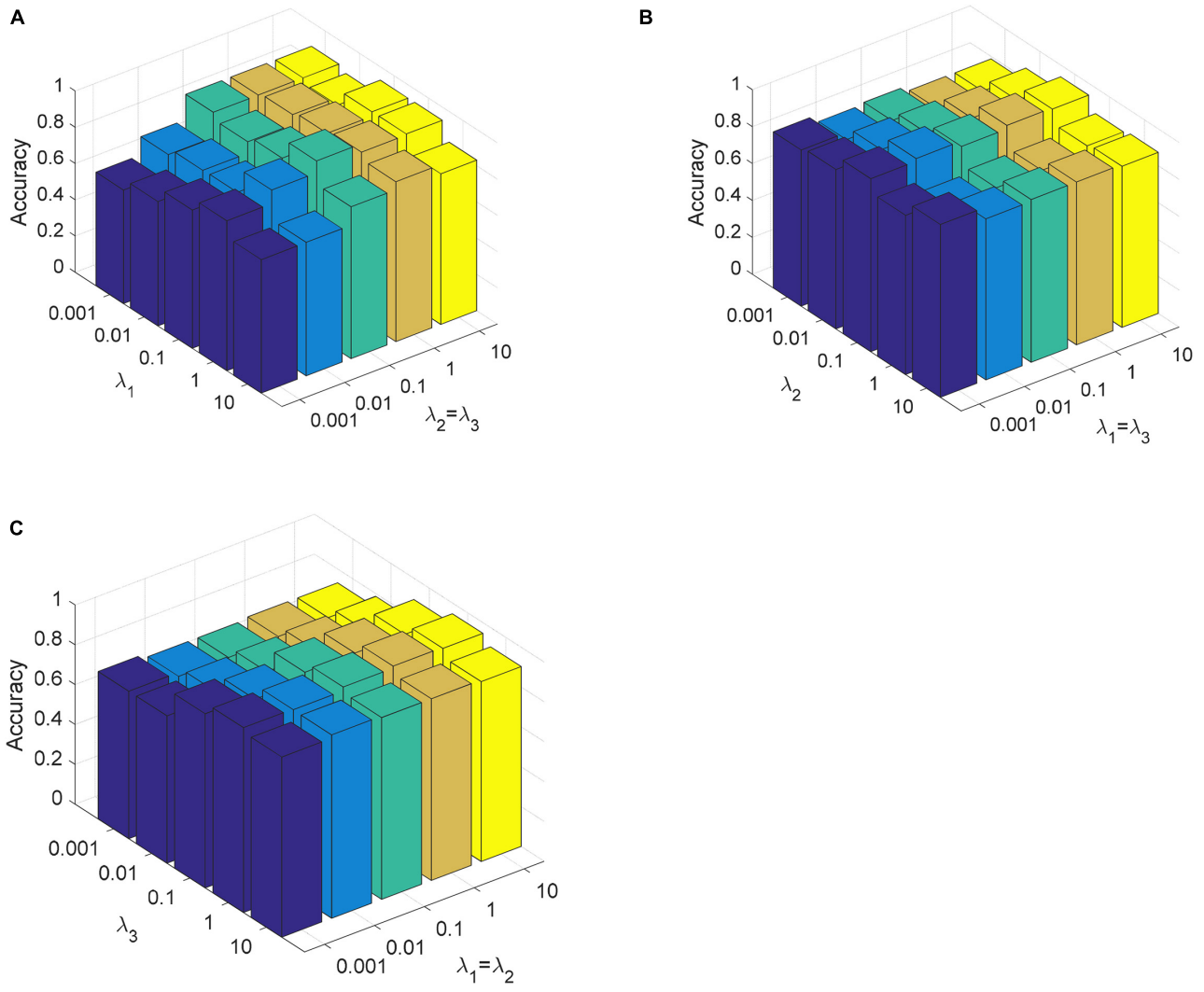
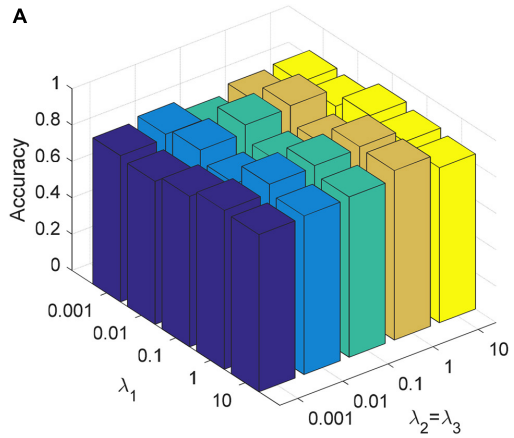
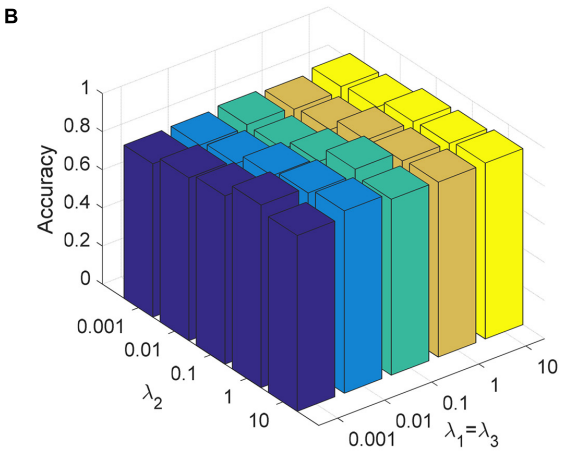


FIGURE 6 | Parameter sensitivity of CDLLC on the Cheng dataset: **(A)** λ_1 and λ_2 (λ_3), **(B)** λ_2 and λ_1 (λ_3), and **(C)** λ_3 and λ_1 (λ_2).





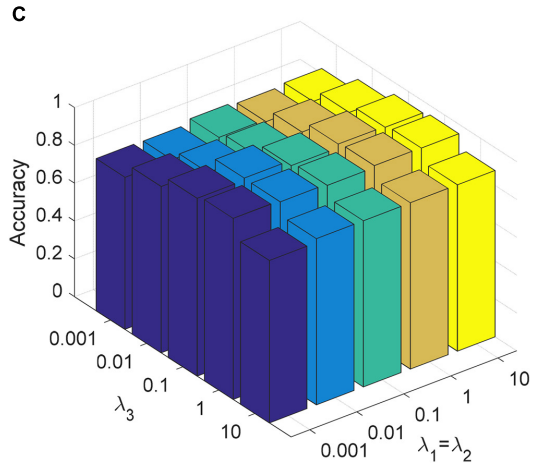
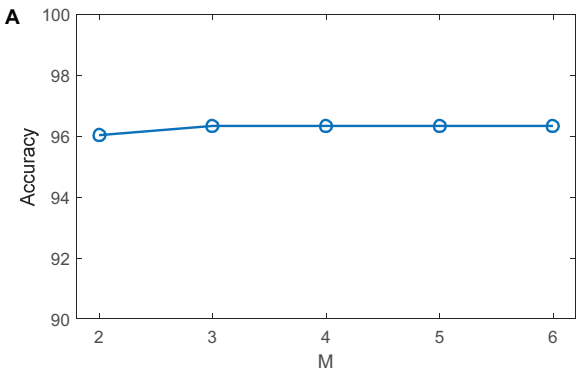


FIGURE 7 | Parameter sensitivity of CDLLC on the REMBRANDT dataset: **(A)** λ_1 and λ_2 (λ_3), **(B)** λ_2 and λ_1 (λ_3), and **(C)** λ_3 and λ_1 (λ_2).



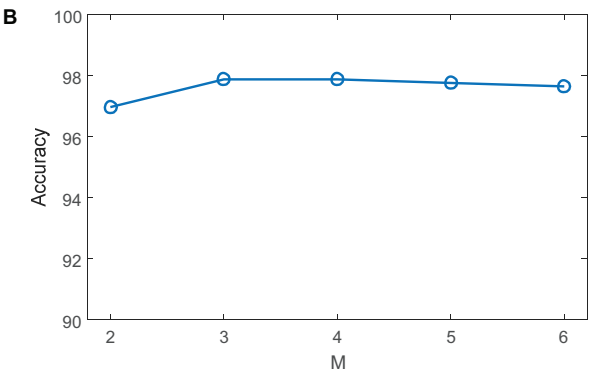


FIGURE 8 | Parameter sensitivity *M* of CDLLC on **(A)** the Cheng dataset and **(B)** the REMBRANDT dataset.

Next, we discuss the number of layers *M* in CDLLC on the Cheng and REMBRANDT datasets. We set *M* in grid {2, 3,..., 6} to evaluate its effect on classification accuracy. The classification result is shown in **Figure 8**. We see that the classification accuracy of CDLLC improves when *M* increases from 1 to 3. When *M* is greater than 3, the classification accuracy of CDLLC is reliable on the Cheng and REMBRANDT datasets.

CONCLUSION

In this study, we propose CDLLC method for brain tumor MR image classification. The CNN structure is utilized to seek sparse representation in the nonlinear space, so that the resulting coding vectors of different classes can give the discriminative approximation. Meanwhile, the proposed method CDLLC uses

the locality constraint of atoms to preserve the manifold structure of the codes. Different from the traditional dictionary learning that uses manual feature extraction, CDLLC extracts the useful CNN features automatically in the architecture of deep learning. Classification of types of meningiomas, gliomas, and pituitary tumors on the Cheng dataset and types of AST, OLI, and GBM on the REMBRANDT dataset is carried out with high performance in accuracy, recall, precision, F1-score, and balance loss. The shortcoming of CDLLC is that the selection of parameters becomes complicated as the number of layers increases. In our next work, we will try to design a more reasonable program to select these parameters. Besides, we will compare various network architectures on CDLLC, such as VGG and GoogLeNet. Also, we will adjust our method so that it could be applied to other medical MR images.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The Cheng dataset analyzed for this study can be found at: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. The REMBRANDT dataset analyzed for this study can be found at: https://wiki.cancerimagingarchive.net/display/Public/REMBRANDT.

AUTHOR CONTRIBUTIONS

TN and XG conceived and developed the theoretical framework of the manuscript. All authors carried out experiment and data process, and drafted the manuscript.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under grant 61806026 and by the Natural Science Foundation of Jiangsu Province under grant BK20180956.

REFERENCES

Adebileje, S. A., Ghasemi, K., Aiyelabegan, H. T., and Saligheh Rad, H. (2017). Accurate classification of brain gliomas by discriminate dictionary learning based on projective dictionary pair learning of proton magnetic resonance spectra. *Magn. Reson. Chem.* 55, 318–322. doi: 10.1002/mrc.4532

Afshar, P., Mohammadi, A., and Plataniotis, K. N. (2018). "Brain tumor type classification via capsule networks," in *in 2018 25th IEEE International Conference on Image Processing (ICIP)*, (New Jersey: IEEE), 3129–3133. doi: 10.1109/ICIP.2018.8451379

Al-Shaikhli, S. D. S., Yang, M. Y., and Rosenhahn, B. (2014). "Coupled dictionary learning for automatic multi-label brain tumor segmentation in flair MRI images," in *In Proceedings of the International Symposium on Visual Computing*, (Cham: Springer), 489–500. doi: 10.1007/978-3-319-14249-4_46

Al-Shaikhli, S. D. S., Yang, M. Y., and Rosenhahn, B. (2016). Brain tumor classification and segmentation using sparse coding and dictionary learning. *Biomed. Tech.* 61, 413–429. doi: 10.1515/bmt-2015-0071

Anaraki, A. K., Ayati, M., and Kazemi, F. (2019). Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybern. Biomed. Eng.* 39, 63–74. doi: 10.1016/j.bbe.2018.10.004

Bunevicius, A., Schregel, K., Sinkus, R., Golby, A., and Patz, S. (2020). Review: MR elastography of brain tumors. *NeuroImage Clin.* 25:102109. doi: 10.1016/j.nicl.2019.102109

Cai, X., Ding, C., Nie, F., and Huang, H. (2013). "On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York: ACM), 1124–1132. doi: 10.1145/2487575.2487701

Chahal, P. K., Pandey, S., and Goel, S. (2020). A survey on brain tumor detection techniques for MR images. *Multimed. Tools Appl.* 79, 21771–21814. doi: 10.1007/s11042-020-08898-3

Chen, X., Nguyen, B. P., Chui, C. K., and Ong, S. H. (2017). An automatic framework for multi-label brain tumor segmentation based on kernel sparse representation. *Acta Polytechnica Hungarica* 14, 25–43. doi: 10.12700/APH.14.1.2017.1.3

Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., et al. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS One* 10:e0140381. doi: 10.1371/journal.pone.0140381

Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., et al. (2016). Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PLoS One* 11:e0157112. doi: 10.1371/journal.pone.0157112

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* 26, 1045–1057. doi: 10.1007/s10278-013-9622-7

Deepak, S., and Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Comput. Biol. Med.* 111:103345. doi: 10.1016/j.compbiomed.2019.103345

Ding, Z., and Fu, Y. (2018). Deep transfer low-rank coding for cross-domain learning. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 1768–1779. doi: 10.1109/TNNLS.2018.2874567

Fujima, N., Homma, A., Harada, T., Shimizu, Y., Tha, K. K., Kano, S., et al. (2019). The utility of MRI histogram and texture analysis for the prediction of histological diagnosis in head and neck malignancies. *Cancer Imaging* 19:5. doi: 10.1186/s40644-019-0193-9

Ge, C., Gu, I. Y. H., Jakola, A. S., and Yang, J. (2020). Deep semi-supervised learning for brain tumor classification. *BMC Med. Imaging* 20:1–11. doi: 10.1186/s12880-020-00485-0

Ghasemi, M., Kelarestaghi, M., Eshghi, F., and Sharifi, A. (2020). T2-FDL: a robust sparse representation method using adaptive type-2 fuzzy dictionary learning for medical image classification. *Expert Syst. Appl.* 158:113500. doi: 10.1016/j.eswa.2020.113500

Ghasemi, M., Kelarestaghi, M., Eshghi, F., and Sharifi, A. (2021). A FDL: a new adaptive fuzzy dictionary learning for medical image classification. *Pattern Anal. Appl.* 24, 145–164. doi: 10.1007/s10044-020-00909-1

Gu, X., Chung, F. L., and Wang, S. (2018). Fast convex-hull vector machine for training on large-scale ncRNA data classification tasks. *Knowl. Based Syst.* 151, 149–164. doi: 10.1016/j.knosys.2018.03.029

Gu, X., Chung, F. L., and Wang, S. (2020a). Extreme vector machine for fast training on large data. *Int. J. Mach. Learn. Cybern.* 11, 33–53. doi: 10.1007/s13042-019-00936-3

Gu, X., Zhang, C., and Ni, T. (2020b). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2020.3006699 [Epub Online ahead of print].

Gumaei, A., Hassan, M. M., Hassan, M. R., Alelaiwi, A., and Fortino, G. (2019). A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. *IEEE Access* 7, 36266–36273. doi: 10.1109/ACCESS.2019.2904145

Jiang, Z., Lin, Z., and Davis, L. S. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105. doi: 10.1145/3065386

Li, H., He, X., Tao, D., Tang, Y., and Wang, R. (2018). Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recognit.* 79, 130–146. doi: 10.1016/j.patcog.2018.02.005

Mittal, M., Goyal, L. M., Kaur, S., Kaur, I., Verma, A., and Hemanth, D. J. (2019). Deep learning based enhanced tumor segmentation approach for MR brain images. *Appl. Soft Comput.* 78, 346–354. doi: 10.1016/j.asoc.2019.02.036

Mohan, G., and Subashini, M. M. (2018). MRI based medical image analysis: survey on brain tumor grade classification. *Biomed. Signal Process. Control* 39, 139–161. doi: 10.1016/j.bspc.2017.07.007

Nandpuru, H. B., Salankar, S. S., and Bora, V. R. (2014). "MRI brain cancer classification using support vector machine," in *Proceedings of the 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*, (New Jersey: IEEE), 1–6. doi: 10.1109/SCEECS.2014.6804439

Ni, T., Gu, X., and Jiang, Y. (2020). *Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy*. Netherland: Springer.

Papyan, V., Romano, Y., Sulam, J., and Elad, M. (2018). Theoretical foundations of deep learning via sparse representations: a multilayer sparse model and its connection to convolutional neural networks. *IEEE Signal Process. Mag.* 35, 72–89. doi: 10.1109/MSP.2018.2820224

Peng, Y., Liu, S., Wang, X., and Wu, X. (2020). Joint local constraint and fisher discrimination based dictionary learning for image classification. *Neurocomputing* 398, 505–519. doi: 10.1016/j.neucom.2019.05.103

Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., and Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J. Comput. Sci.* 30, 174–182. doi: 10.1016/j.jocs.2018.12.003

Song, J., Xie, X., Shi, G., and Dong, W. (2019). Multi-layer discriminative dictionary learning with locality constraint for image classification. *Pattern Recognit.* 91, 135–146. doi: 10.1016/j.patcog.2019.02.018

Sulam, J., Papyan, V., Romano, Y., and Elad, M. (2018). Multilayer convolutional sparse modeling: Pursuit and dictionary learning. *IEEE Trans. Signal Process.* 66, 4090–4104. doi: 10.1109/TSP.2018.2846226

Sun, R., Wang, K., Guo, L., Yang, C., Chen, J., Ti, Y., et al. (2019). A potential field segmentation based method for tumor segmentation on multi-parametric MRI of glioma cancer patients. *BMC Med. Imaging* 19:48. doi: 10.1186/s12880-019-0348-y

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, 1–41. doi: 10.3322/caac.21660

Thirumurugan, P., and Shanthakumar, P. (2016). Brain tumor detection and diagnosis using ANFIS classifier. *Int. J. Imaging Syst. Technol.* 26, 157–162. doi: 10.1007/978-3-642-31552-7_35

Tong, J., Zhao, Y., Zhang, P., Chen, L., and Jiang, L. (2019). MRI brain tumor segmentation based on texture features and kernel sparse coding. *Biomed. Signal Process. Control* 47, 387–392. doi: 10.1016/j.bspc.2018.06.001

Vu, T. H., Mousavi, H. S., Monga, V., Rao, G., and Rao, U. A. (2015). Histopathological image classification using discriminative feature-oriented dictionary learning. *IEEE Trans. Med. Imaging* 35, 738–751. doi: 10.1109/TMI.2015.2493530

Wang, J., Kwon, S., and Shim, B. (2012). Generalized orthogonal matching pursuit. *IEEE Trans. Signal Process.* 60, 6202–6216. doi: 10.1109/TSP.2012.2218810

Wu, G., Chen, Y., Wang, Y., Yu, J., Lv, X., Ju, X., et al. (2017). Sparse representation-based radiomics for the diagnosis of brain tumors. *IEEE Trans. Med. Imaging* 37, 893–905. doi: 10.1109/TMI.2017.2776967

Yang, Y., Yan, L. F., Zhang, X., Han, Y., Nan, H. Y., Hu, Y. C., et al. (2018). Glioma grading on conventional MR images: a deep learning study with transfer learning. *Front. Neurosci.* 12:804. doi: 10.3389/fnins.2018.00804

Zeng, K., Zheng, H., Cai, C., Yang, Y., Zhang, K., and Chen, Z. (2018). Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Comput. Biol. Med.* 99, 133–141. doi: 10.1016/j.compbiomed.2018.06.010

Zhang, Y., Chen, C., Tian, Z., Feng, R., Cheng, Y., and Xu, J. (2019). The Diagnostic Value of MRI-Based Texture Analysis in Discrimination of Tumors Located in Posterior Fossa: a Preliminary Study. *Front. Neurosci.* 13:1113. doi: 10.3389/fnins.2019.01113

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 02 June 2021
doi: 10.3389/fnins.2021.638976

Multimodal Medical Supervised Image Fusion Method by CNN

Yi Li[1,2], Junli Zhao[1], Zhihan Lv[1] and Zhenkuan Pan[3]**

[1] College of Data Science Software Engineering, Qingdao University, Qingdao, China, [2] Business School, Qingdao University, Qingdao, China, [3] College of Computer Science and Technology, Qingdao University, Qingdao, China

This article proposes a multimode medical image fusion with CNN and supervised learning, in order to solve the problem of practical medical diagnosis. It can implement different types of multimodal medical image fusion problems in batch processing mode and can effectively overcome the problem that traditional fusion problems that can only be solved by single and single image fusion. To a certain extent, it greatly improves the fusion effect, image detail clarity, and time efficiency in a new method. The experimental results indicate that the proposed method exhibits state-of-the-art fusion performance in terms of visual quality and a variety of quantitative evaluation criteria. Its medical diagnostic background is wide.

Keywords: deep learning, image fusion, CNN, multi-modal medical image, medical diagnostic

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Xiaoqing Gu,
Changzhou University, China
Juan Yang,
Suzhou University, China

****Correspondence:***
Yi Li
lyqgx@126.com
Zhenkuan Pan
Panzk@qdu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *08 December 2020*
Accepted: *02 March 2021*
Published: *02 June 2021*

Citation:
Li Y, Zhao J, Lv Z and Pan Z (2021) Multimodal Medical Supervised Image Fusion Method by CNN. Front. Neurosci. 15:638976. doi: 10.3389/fnins.2021.638976

INTRODUCTION

With the continuous recommendation of medical image processing research in recent years, image fusion is an effective solution that automatically detects the information in different images and integrates them to produce one composite image in which all objects of interest are clear. Image fusion (Zhang et al., 2018; Liu et al., 2019; Ma et al., 2019; Pan and Shen, 2019) is a specific algorithm to combine two or more images into a new image. Because of its wide application value, multimodal medical image fusion is an important branch in the field of image fusion. Because of the wide use of multimodal medical images (Yizhang et al., 2019, 2021), this problem has become a hot topic in recent years. At present, in the field of image fusion, deep learning method is a representative method, and it is also one of the research focuses in recent years. Many scholars at home and abroad have conducted on deep learning research and widely applied the research results in image processing and other fields. The recent research in image fusion based on deep learning is as follows: pixel-level image fusion, convolutional neural networks (CNNs) (Liu et al., 2017; Cheng et al., 2018; Hou et al., 2018; Schlemper et al., 2018; Shao and Cai, 2018; Sun et al., 2019), convolutional sparse representation (Li and Wu, 2019), stacked autoencoders (Ahmad et al., 2017; Bao et al., 2017; Jiao et al., 2018; Ma et al., 2018; Chen et al., 2019; Kosiorek et al., 2019), and deep belief network (DBN) (Dong et al., 2016; Chen and Li, 2017; Ye et al., 2018). As an example in CNN research, Shao and Cai (2018) propose a remote sensing image fusion method based on CNN. Cheng et al. (2018) focused on challenges within-class diversity and between-class similarity. The proposed D-CNN models are trained by optimizing a new discriminative objective function. A metric learning regularization term on the CNN features is imposed in this research in order to outperform the existing baseline methods and achieve state-of-the-art results. Level measurement and fusion rule are still important problem in CNN fusion modal. Emphasis in the work of Liu et al. (2017) is laid on these two aspects of research. It proposes a new multifocus image fusion method to overcome the difficulty faced by the existing fusion methods. The method mainly consists of two procedures. First, spatial and

spectral features are, respectively, extracted by convolutional layers with different depth. Second, the extracted features from the former step are utilized to yield fused images. In addition, CNN also has a more in-depth study in the field of image action recognition research, image reconstruction. An effective method (Hou et al., 2018) to encode the spatiotemporal information of a skeleton sequence into color texture images is proposed, referred to as skeleton optical spectra, and employs CNNs (ConvNets) to learn the discriminative features for action recognition. This is a typical result of action recognition research. Then, a framework for reconstructing dynamic sequences of two-dimensional cardiac magnetic resonance (MR) images from undersampled data using a deep cascade of CNNs to accelerate the data acquisition process is proposed in research (Schlemper et al., 2018). In addition to the above results, CNN can also be widely used in the study of image super-resolution. In the study of Sun et al. (2019), a novel parallel support vector mechanism (SVM)–based fusion strategy to take full use of deep features at different scales as extracted by the MCNN:Multi-task convolutional neural network structure is proposed. An MCNN structure with different sizes of input patches and kernels is designed to learn multiscale deep features. After that, features at different scales were individually fed into different support vector machine (SVM) classifiers to produce rule images for preclassification. A decision fusion strategy is then applied on the preclassification results based on another SVM classifier. Finally, superpixels are applied to refine the boundary of the fused results using region-based maximum voting.

Furthermore, the more recent theoretical research in the field of deep learning is as follows:

In recent years, deep learning is also widely used in image super-resolution process. In this research, Huang et al.'s 2018 research focused on two defect problems in direction-of-arrival and multiple input multiple output. Deep learning can be applied to solve this problem that resource allocation issue is an obstacle (Takaishi et al., 2018). Importantly, brief advantages of the deep learning based communication schemes are demonstrated in this work. Furthermore, in mobile video research, Liu et al. (2018) point out that the problem of learning temporal dynamics from two aspects is tackled. It is focused on research in the complexity degree of motion among neighboring frames using spatial alignment networks.

The vast majority of these studies focus on the study of single images; the studies of multiple images have been rarely involved. But medical images have specific practical requirements, information richness, and high clarity. Image fusion can increase the amount of information in a single image. To solve this practical medical problem, we propose the method of completion in fusion and hyperscore simultaneously. That is why we are doing this. Multimodal image is a major category in medical image.

In these documents, the requirements of multimodal images for information and clarity have been repeatedly emphasized. In order to effectively meet the needs of the aforementioned medical images and make tentative research on the development of automatic diagnostic technology, supervised deep learning methods were used to achieve image fusion. In this article, deep learning model is intended to be introduced into the field of image fusion. It is intended to develop a new idea of image fusion based on supervised deep learning. We can obtain a new model through the establishment of image training databases in successful fusion results. It is suitable for image fusion to improve the efficiency and accuracy of image process.

CNN

Deep learning is a hot topic in recent years. Many scholars have focused on their research in several common models such as AutoEncoder (AE), CNN, Restricted Boltzmann Machine (RBM), and DBN. Next, this article summarizes these three models as follows:

As you can see, the image of the ship on the far left is our input layer, which the computer understands as a number of matrices, which is basically the same as DNN. Then, there is the convolution layer, which is specific to CNN, which will be discussed later in (**Figure 1**). The convolution layer activation function uses a ReLU. We introduced the RELU activation function in DNN, which is very simple ReLU(x) = max(0,x). Behind the convolution layer is the pooling layer, which is also specific to CNN and will be covered later. Note that there is no activation function for the pooled layer. The convolution formula commonly used here is shown in Formula (1). The detailed convolution process is shown in **Figure 2**. The third step of the model is to complete the pooling.

$$s(i,j) = (X^* W)(i,j) + b = \sum_{k=1}^{n_in} (X_k {}^* W_k)(i,j) + b \qquad (1)$$

FRAMES AND ALGORITHMS

The proposed framework is shown in **Figure 3**. It consists of two major steps: model learning and fusion test. In the first step, the parameters in the DBN model are learned by training the multiple groups of images in the train datasets. Registration processing and pixel alignment in these train images have been achieved in advance. In the second fusion test process, the multiple groups of test images are entered into the model that the learning and training have achieved, and then the fusion process is completed. Next, the final synthesis obtains the fused image.

The new method proposed in this article can realize the fusion task of multimodal medical images.

The specific implementation process is as follows:

Training process:

- In this article, noise removal, registration, standardization, and other preprocessing work will be carried out for a large number of images in datasets. These images will be divided into training datasets and testing datasets in the same level according to the standards of depth learning model.
- The image block size is determined on the training dataset and the test dataset image, respectively. On the two datasets, we use the same block size to complete the model calculation. The size is determined by the standard and the result of the main reference image fusion. If it needs a high

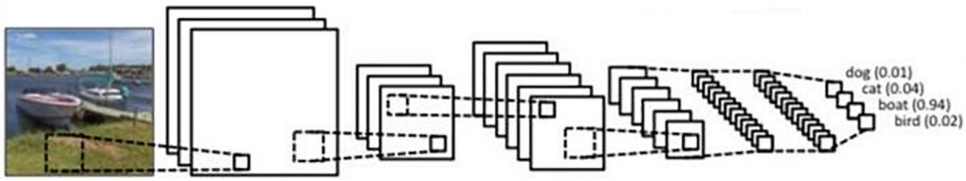


FIGURE 1 | Training process of CNN.

precision, it must be selected smaller. If it needs a higher calculation speed, it can be selected larger. In order to better meet the needs of medical diagnosis, different sizes of the calculation model can be used alternately.

- After the training data collection is completed, a fusion calculation model is generated, and related parameters are further optimized. Parameter optimization can be achieved by using automatic optimization.

Testing process:

The final step is also a key step in the process of model testing, in two separate cases. First, the two sets of multimodal images were entered into the fusion model to complete the fusion results. Next, batch processing in multiple group images are achieved. The fusion results were obtained by refactoring, and then these results were output. This model can also complete the process of multi-image fusion at the same time. The specific process is as follows:

- Preprocessed multi-images
- To be fused as input
- Image mapping
- Feature extraction and other work
- Get the fusion results
- Image reconstruction
- To get a fused image

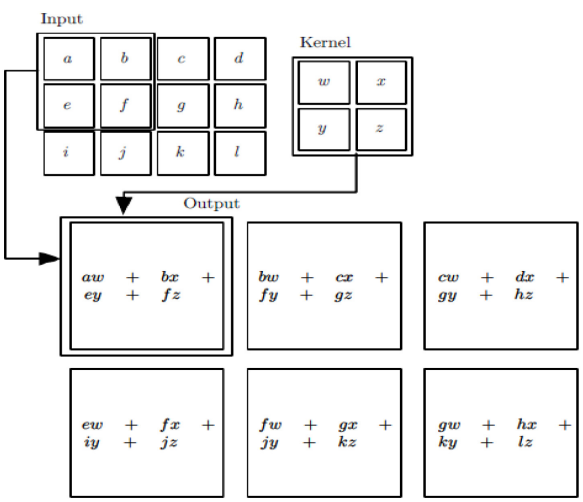


FIGURE 2 | Convolution process.

EXPERIMENTAL AND ANALYSIS

Extensive experiments are conducted on different medical image pairs in this part, e.g., computed tomography (CT), MR imaging (MRI), and single-photon emission CT (SPECT). The experimental images are all from public data[1]. The evaluation metrics used in this article are EOG,RMSE, PSNR, ENT, SF, REL, SSIM, H, MI, MEAN, STD, GRAD, Q_0, Q_E, Q_W and $Q^{AB/F}$; readers can refer to References (Wang and Bovik, 2002; Piella and Heijmans, 2003; Yin et al., 2013; Ma et al., 2015; Liang et al., 2016) for more details. The ranges of Q_0, Q_E, Q_W and $Q^{AB/F}$ are in [0,1], and the larger the value, the better the fused result. To reduce variation, each experiment is repeated more than 100 times, and their mean values are recorded.

Databases of Learning

Among the extensive multimodal medical images, the classic images can be divided into two categories: MRI images and CT images. MRI images are more accurate, and its information is more abundant and accurate, especially for human tissue structure and details. CT images provide rich anatomical structure images of the human body. Clinically, images of various anatomical structures can be observed through bone windows, soft tissue windows, and lung windows, and the details of organs can be reflected in detail from a certain angle. SPECT images are one of the typical image types in CT imaging technology. Therefore, the construction of the datasets in the experiment we chose is composited with more classical MRI, CT, and SPECT image in the multimode image. The images were all derived from the standard medical image database and were registered before the experiment. Typical part sets of images are selected to display in this article, as shown in **Figures 4**, **5**. The first image in **Figures 4**, **5** is an MRI image, the second image is SPECT image, and the third image is a target image.

(1) CT and MRI images data
(2) MRI and SPECT images data

Experimental Results on CT and SPECT Images

Experimental Results of the Proposed Method

We have conducted detailed experimental studies on the different disturbance levels of the test images. Here we show the typical two sets of experimental results in the article, as shown in

[1]http://www.metapix.de/download.htm#opennewwindow

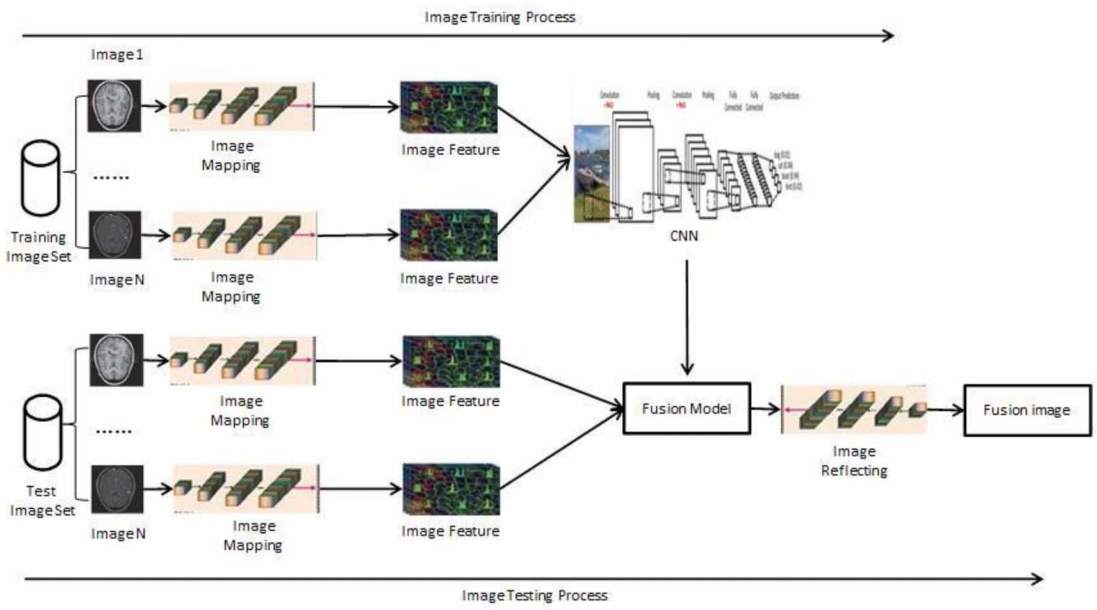


FIGURE 3 | Model of supervised image fusion based on CNN.

Figures 6, **7** and **Tables 1**, **2**. The perturbation parameters we use are randomly generated; here, we simply refer to "primary disturbance" and "secondary disturbance." The first set of experimental results in **Figure 8** is the result of testing in the case of a primary disturbance, and the second set of experimental results is the result of testing in the case of a secondary disturbance.

Comparison With Other Classical Image Fusion Methods Based on Deep Learning

In the experiment, we also carried out this method and compared with the "Li" method in literature (Li and Wu, 2019) and the "Liu" method in literature. The parameters in the model can are obtained by learning, they cannot be determined previously. We have conducted similar experiments on multiple sets of images in the datasets. Here, we present a representative set of experimental results in the article, as shown in **Figure 7**.

Analysis of Experimental Results

Visual Quality of the Fused Images

First, the visual quality of the fused images obtained by the proposed method is better than the other methods. The fused images obtained by our method look more natural. They produce sharper edges and higher resolution. In addition, the detailed information and interested features are better preserved to some extent, as shown in **Figures 6**, **7** and **Tables 1**, **2**.

In particular, from the area circled by the red calibration frame, it can be seen that the fusion results are clear, the edges are clear, the information is clearly contrasted, the contrast is obvious, and the key information in the image can be reflected, and the virtual shadow is effectively removed. Moreover, it is clear that the information contained in the fusion image already covers most of the information in the two multimodal images of CT and MRI, MRI and SPECT to be fused, it can be effectively supplementing the deficiencies of the single MRI/CT/SPECT image information. The increase in the amount of information brings changes to the improvement of medical imaging diagnosis undoubtedly. More valuable information can be used to support effective diagnosis. It also brings possibilities for the study of "automatic diagnosis" technology and conducts tentative research. The result of tests 1–10 shown in **Figure 6** is the key main similarity between the fusion result and the target image. It can be reflected that the result of the method fusion in this article has covered the main key information of the target image in a more comprehensive way. This can explain the validity of this method further.

Analysis of Evaluation Data

From the evaluation data shown in **Figure 8**, we can find the following:

For the first experiment, the proposed method achieves excellent results when using evaluation metrics :EOG, RMSE, PSNR, ENT, SF, REL, SSIM, H, MI, MEAN, STD, GRAD, Q_0, Q_E, Q_W, and $Q^{AB/F}$. The results show that our methods are effective in image fusion.

Specifically, the result in tests 1–8 of CT and MRI is somewhat better than SR with respect to Q_W.

For tests 2–10 experiment results of MRI and SPECT, the proposed method yields outstanding results in terms of Q_W, EOG, RESE, PSNR, and REL.

Hence, our method can realize the image fusion task and capture the details of the image compared to other fusion methods.

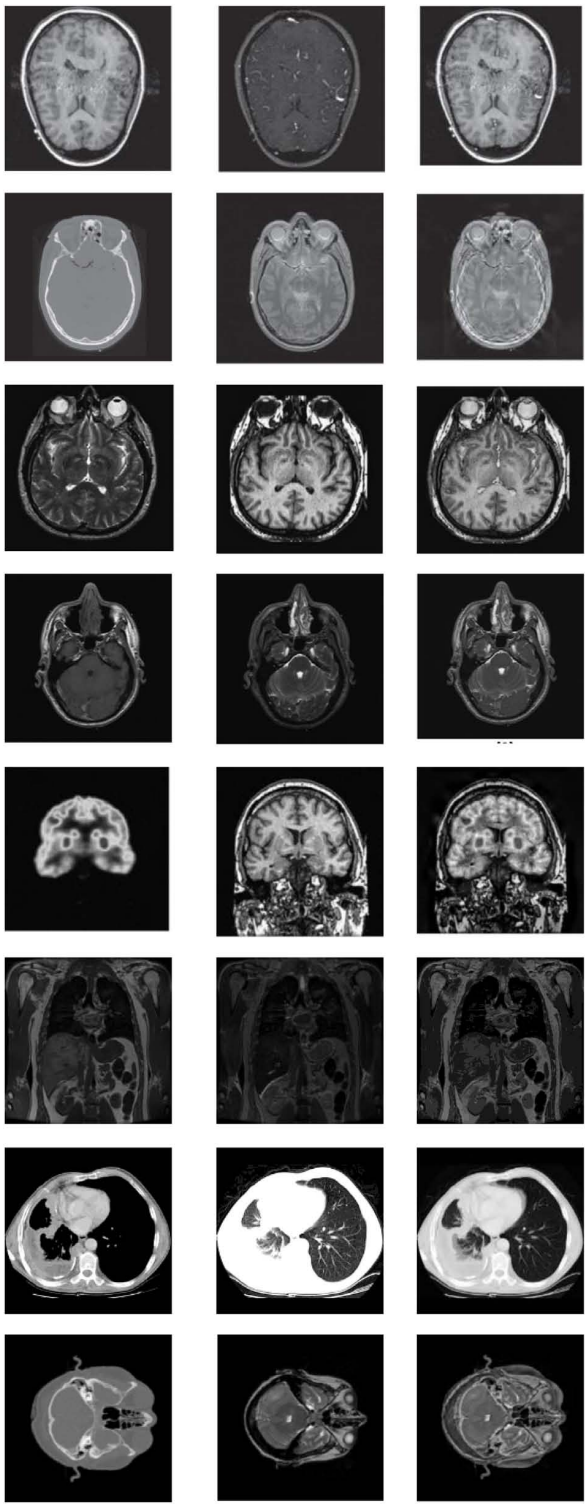


FIGURE 4 | Databases of learning images.

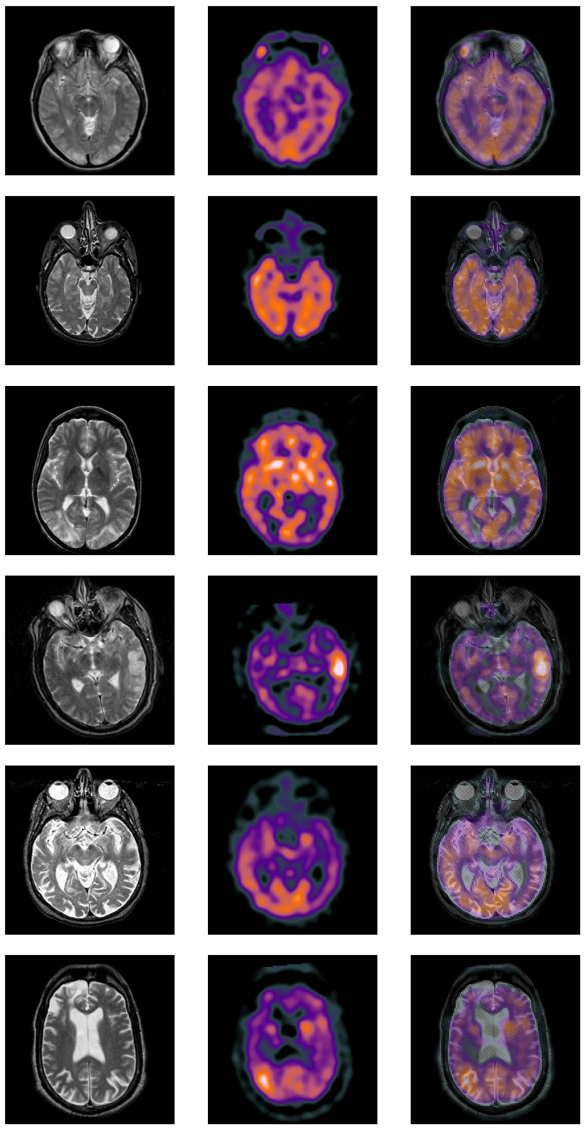


FIGURE 5 | Databases of learning images.

They results also verify that the process of learning, testing, fusion in this method not only can introduce fusion effectively, but also can make the important information and details in the images integrated by the depth learning model prominently. Fusion result can be covered in key information in two types of medical images, and it can obtain more informative and complete medical image information. However, it is slightly inferior in terms of REL and STD, indicating that the fusion process will cause a loss of information; how to learn model parameters, increase the size of the training data, and improve the degree of training accuracy are our issues for further study.

Correlation Between Grid Block Size and Method Validity

In this article, when computing the model learning, we consider small, overlapping image patches instead of direct whole images. In addition, each image is divided into small blocks of size $n \times n$. Thus, in this section, we intend to explore how the size

of the image path affects the fusion performance. Four pairs of source images used in the previous section are utilized in this experiment, and the image is divided into small blocks of size $2i(i = 0, 1, 2, 3\ldots)$, in order not to end up with extra grid blocks smaller than the intended size. The larger the value of i, the larger the size of the grid block we obtained. For comparison, evaluation metrics EOG, RMSE, PSNR, ENT, SF, REL, SSIM, H, MI, MEAN, STD, GRAD, $_{Q_0}$, $_{Q_E}$, $_{Q_W}$, and $_{Q^{AB/F}}$ are used, and the curves of experimental results are plotted and depicted in **Figure 9**. The results of the experiment are marked with the mean of 50 operating results in the same environment. The typical six experimental results are presented in the article in **Figure 6**. We have conducted similar experiments and analyses on other undisplayed indicators and have similar conclusions.

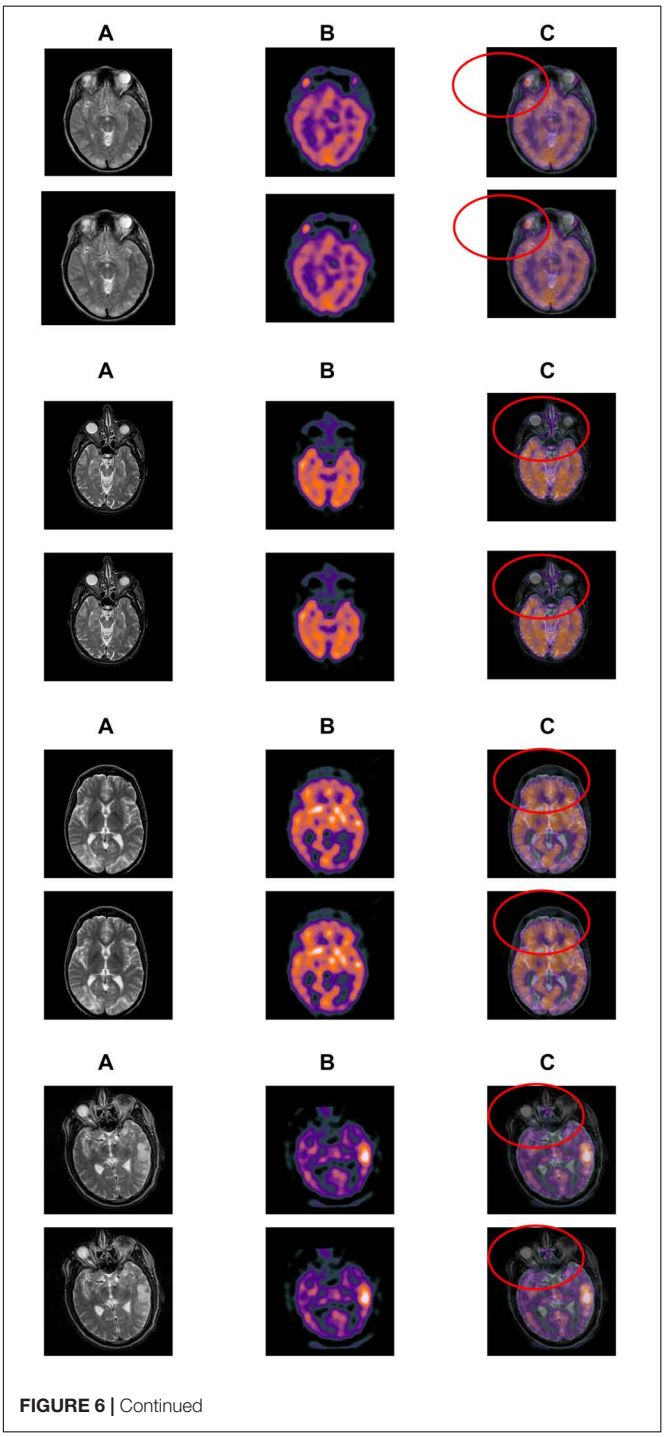


FIGURE 6 | Continued

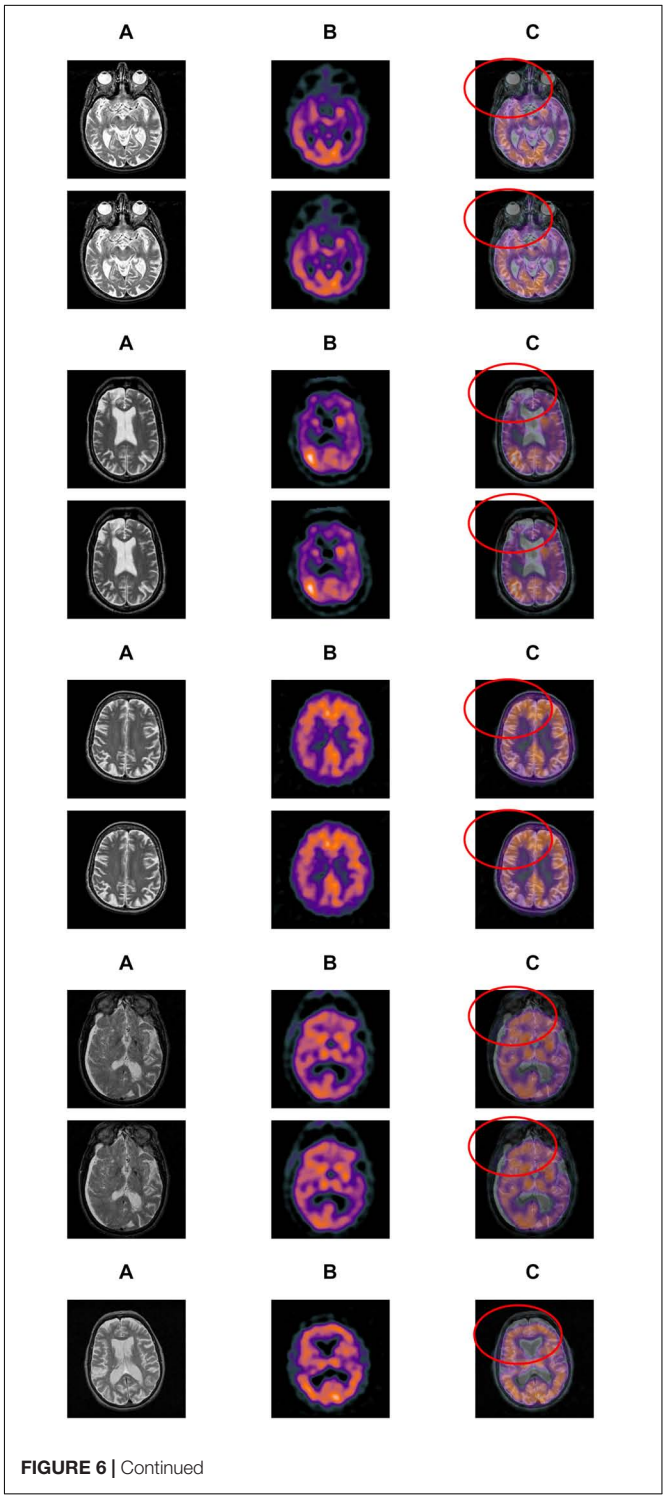


FIGURE 6 | Continued

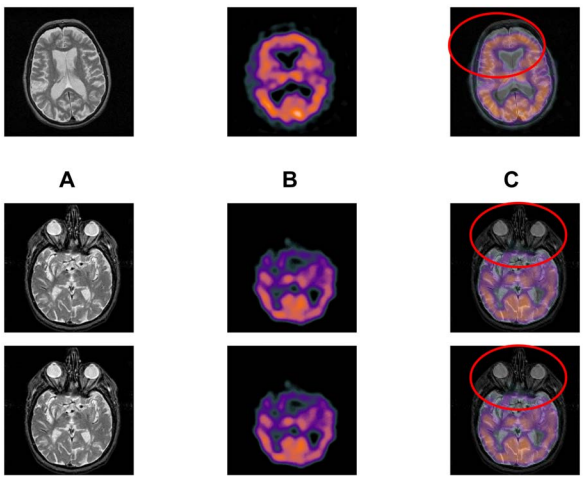


FIGURE 6 | Experiment results on images: **(A)** original image—A; **(B)** original image—B; **(C)** experiment results of image fusion.

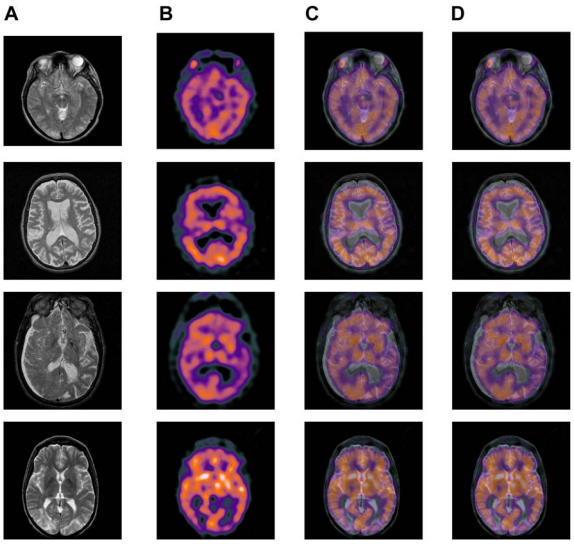


FIGURE 7 | Experiment results on multifocus images: **(A)** original image—A; **(B)** original image—B; **(C)** Liu; (d) Li.

In RMSE, we have conducted detailed experimental studies on the values of $i(i = 1, 2, 3, 4)$, respectively. Through repeated studies of experimental data, we found that when i takes 1 and 2, that is, when the block size takes 2×2 and 4×4, the error rate of the fusion result is relatively low. When the value i gradually climbed to 3 and 4, the error rate increased significantly; the curve gradient increased significantly. This phenomenon has similar conclusions for the 10 sets of images used in testing. It can be explained that selecting different i values to achieve the best fusion effect is still different for different fusion images. How to use automatic optimization to dynamically determine the value of i need further in-depth research. In terms of image entropy, we also conducted detailed experimental studies on the values of $2i(i = 1, 2, 3, 4)$, respectively. Through repeated studies of experimental data, we found that when i takes 1 and 2, the information entropy of the fusion results is relatively high. That is to say, the information contained in the fusion image at this time is richer than the result that the block size increases. With the continuous growth of the block size, when the value i gradually climbed to 3 or 4, the information entropy was significantly reduced, and along with the increase of i, its decline speed increased significantly, and the curve gradient increased significantly. This phenomenon has reached similar conclusions for the 10 groups of images. In particular, in the "test 8" group image experiment, when the value of i increased from 2 to 3, the information entropy of the corresponding fusion image accelerated significantly. Similar conclusions can be obtained on the MEAN and GRAD. In the MEAN analysis results, when the value of similar groups of data gradually rises to 3 or 4, the mean decreases significantly.

In the experimental data analysis results of REL and SSIM, we found that there was a difference with the previous four experimental results. In these two indicators, along with the increase of i, the REL and SSIM indicators of the fusion image do not appear decline trend significantly, but only slowly slid. It shows that the fusion results obtained by using our fusion method are relatively stable. With the change of the size of the blocks, there is only a small decline. From another aspect, it also shows that our proposed method has a certain degree of stability.

TABLE 1 | Comparison of results.

Indices	Indices											
Test images no.	RMSE	PSNR	Entropy	REL	SSIM	H	MI	Mean	STD	ENT	GRAD	Time (s)
Test 1	0.3695	8.6466	0.8918	0.2014	0.9943	0.1987	0.0580	0.0309	0.1730	0.1987	0.0511	1.2776
Test 2	0.3273	9.7007	0.9805	0.4291	0.9955	0.4939	0.1404	0.1080	0.3104	0.4939	0.1718	1.8766
Test 3	0.3576	8.9332	1.1560	0.4016	0.9947	0.5146	0.1483	0.1149	0.3189	0.5146	0.1797	1.9045
Test 4	0.4786	6.4112	2.8396	0.2986	0.9907	0.8361	0.1125	0.2663	0.4420	0.8361	0.3562	1.7547
Test 5	0.4027	7.9010	1.7262	0.3904	0.9932	0.6382	0.1670	0.1617	0.3681	0.6382	0.2093	1.9323
Test 6	0.4113	7.7176	1.0713	0.2315	0.9929	0.3335	0.0893	0.0615	0.2403	0.3335	0.1006	1.5789
Test 7	0.3710	8.6134	0.9708	0.2967	0.9943	0.3194	0.1109	0.0580	0.2337	0.3194	0.0872	1.4625
Test 8	0.4505	6.9269	2.5827	0.3535	0.9916	0.8121	0.1374	0.2505	0.4333	0.8121	0.3433	1.9378
Test 9	0.5499	5.1947	3.1892	0.1770	0.9878	0.9378	0.1259	0.3542	0.4783	0.9378	0.4771	1.6932
Test 10	0.4246	7.4413	1.2606	0.2879	0.9926	0.6866	0.0880	0.1830	0.3867	0.6866	0.2644	1.8334

TABLE 2 | Comparison of results.

Indices	CT and MRI				MRI and SPECT			
Test images no.	Q_0	Q_E	Q_W	$Q^{AB/F}$	Q_0	Q_E	Q_W	$Q^{AB/F}$
Test 1	0.9886	0.8002	0.8707	0.8911	0.7658	0.1177	0.5633	0.2164
Test 2	0.7406	0.3092	0.7411	0.5919	0.7518	0.1578	0.5928	0.2431
Test 3	0.5953	0.1622	0.4622	0.4257	0.7629	0.1363	0.5581	0.2239
Test 4	0.8515	0.3678	0.8870	0.5432	0.6145	0.2247	0.6470	0.2701
Test 5	0.6273	0.2814	0.6935	0.3949	0.6692	0.1350	0.5861	0.2359
Test 6	0.8322	0.5549	0.8716	0.7103	0.7403	0.1392	0.5645	0.2381
Test 7	0.8392	0.3273	0.8057	0.6090	0.7702	0.1511	0.5670	0.2276
Test 8	0.7258	0.3213	0.7667	0.5359	0.6948	0.2668	0.6100	0.2897
Test 9	0.4821	0.2439	0.6412	0.3240	0.4683	0.1519	0.6474	0.3281
Test 10	0.7572	0.2698	0.6946	0.3338	0.5972	0.2298	0.6886	0.2961

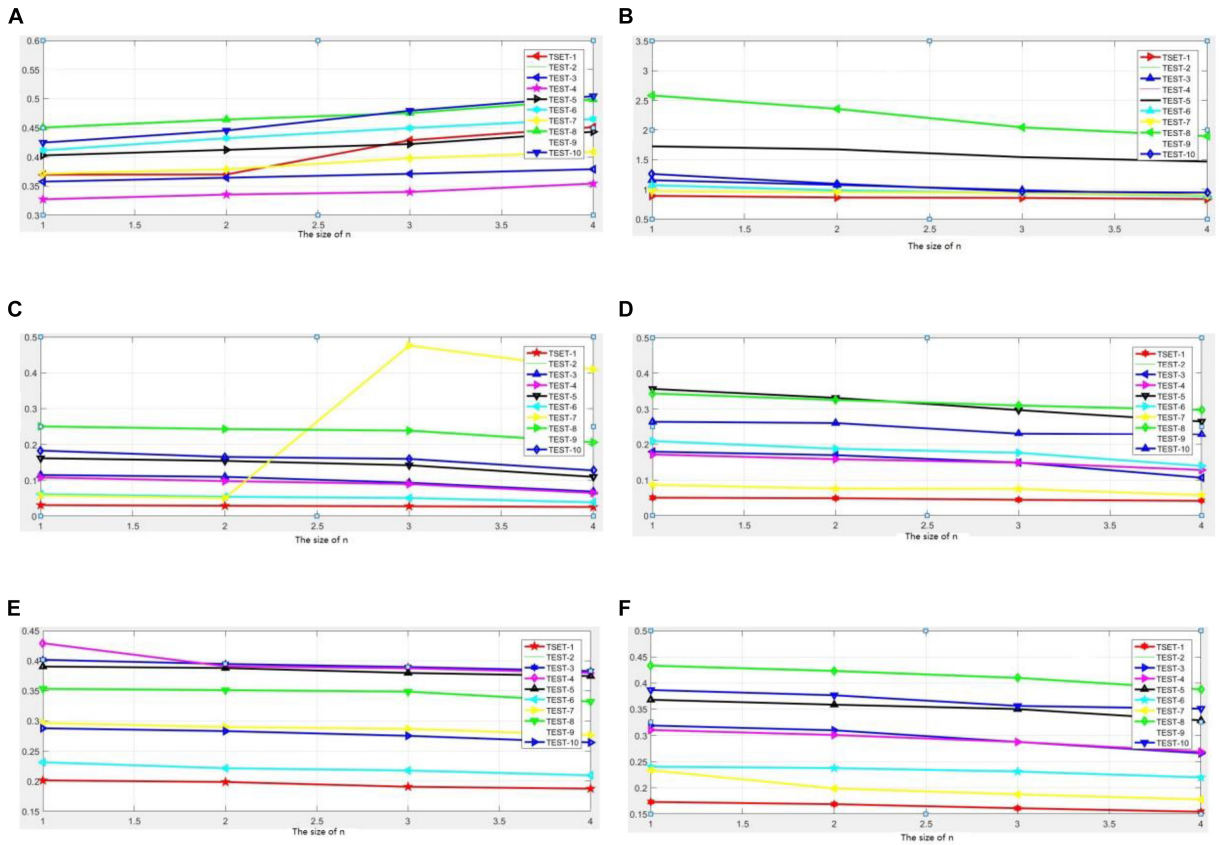


FIGURE 8 | Comparison in block size and fusion performance of the results: **(A)** RMSE, **(B)** entropy, **(C)** MEAN, **(D)** GRAD, **(E)** REL, **(F)** STD.

In summary, the determination of the size of the block has a certain influence on the quality of the fusion result, and the small value of *i* will increase the computational workload and time costs inevitably. The better fusion effect can be obtained; this conclusion can be seen from the aspects of error rate, information entropy, mean value, gradient, similarity, etc. Along with the value of *i* continues to grow, the quality of fusion image shows a certain decline in the above indicators, especially *i* takes in Liu et al. (2019), Pan and Shen (2019), this change is even more obvious. This shows that the accuracy of the calculation has decreased with the change of value *i*. As opposed to this, the time cost has decreased. As far as the REL and SSIM similarity metrics are concerned, they are not so sensitive to the changes in the value of *i*, indicating that our proposed method has a certain degree of stability, and this method can be applied to multimodal images effectively. That raises another question that the exact number of *i* that is taken is more appropriate. The exact number of *i* can depend on the specific image fusion problem. The more emphasis on the accuracy of the calculation, the smaller it can be obtained. Conversely, it can be moderately enlarged in order to improve the efficiency of the operation.

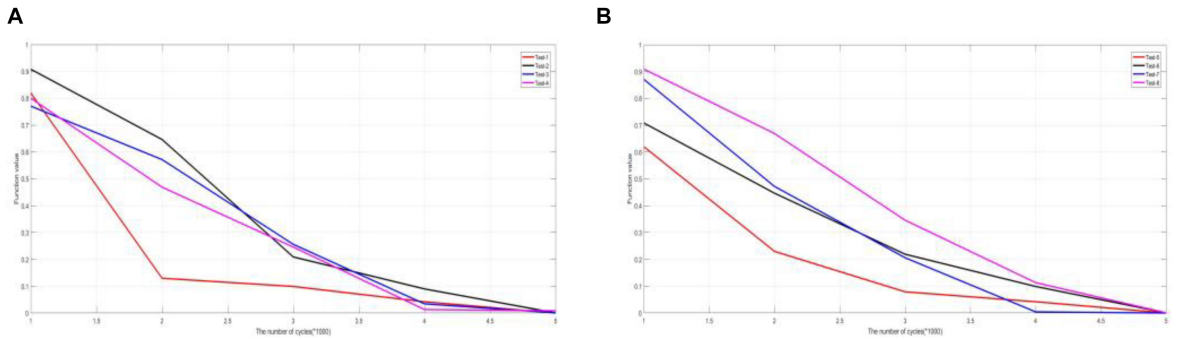


FIGURE 9 | Comparison of experimental results in convergence **(A)** test image 1–4, **(B)** test image 5–8.

Method Convergence Analysis

In this section, we take an experimental study on 10 groups of CT and MRI and 10 groups of MRI and SPECT images. The curves of experimental results are plotted and depicted in **Figure 9**. The results of the experiment are marked with the mean of 50 operating results in the same environment. The typical six experimental results are presented in the article in **Figure 9**. We have conducted similar experiments and analyses on other undisplayed indicators and have similar conclusions.

A large number of experimental data show that after the number of cycles n exceeds 5,000, the function values converge to 0. Function convergence further shows that the method has stability, and there will be no difference in the results in a large number of experiments.

This conclusion can be guaranteed to be accurate and effective in applying this method to medical diagnosis. It is possible that this method is applied in practical medical image processing environment. It also provides an opportunity for further research on the method.

CONCLUSION

The application of deep learning techniques to multimodal medical image has been proposed in this article. This article reviews the recent advances achieved in DL image fusion and puts forward some prospects for future study in the field. The primary contributions of this work can be summarized as the following three points.

Deep learning models can extract the most effective features automatically from data to overcome the difficulty of manual design. The method in this article can achieve the multimodal medical image fusion by CNN. Experimental results indicate that the proposed method exhibits state-of-the-art fusion performance in terms of visual quality and a variety of quantitative evaluation criteria. This proposed method can better adapt to the actual needs of medical multimodal image fusion.

In conclusion, the recent research achieved in DL image fusion and super-resolution exhibits a promising trend in the field of image fusion with a huge potential for future improvement. It is highly expected that more related researches would continue in the coming years to promote the development of image fusion.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

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.

AUTHOR CONTRIBUTIONS

YL was researcher in image processing and pattern recognition, with a major in mathematics and computer science, having plenty of work experience in virtual reality and augmented reality projects, was engaged in the application of computer medical image diagnosis, and her research application fields range widely from deep research fields to everyday lives. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grants 61702293, 11472144, ZR2019LZH002, and 61902203).

ACKNOWLEDGMENTS

We first sincerely thank the editors and reviewers for their constructive comments and suggestions, which are of great value to us. We would also like to thank Xiaojun Wu from Jiangnan University.

REFERENCES

Ahmad, M., Yang, J., Ai, D., Qadri, S. F., and Wang, Y. (2017). "Deep-stacked auto encoder for liver segmentation," in *Proceedings of the Chinese Conference on Image and Graphics Technologies* (Singapore: Springer), 243–251. doi: 10.1007/978-981-10-7389-2_24

Bao, W., Yue, J., and Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One* 12:e0180944. doi: 10.1371/journal.pone.0180944

Chen, H., Jiao, L., Liang, M., Liu, F., Yang, S., and Hou, B. (2019). Fast unsupervised deep fusion network for change detection of multitemporal SAR images. *Neurocomputing* 332, 56–70. doi: 10.1016/j.neucom.2018.11.077

Chen, Z., and Li, W. (2017). Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.* 66, 1693–1702. doi: 10.1109/tim.2017.2669947

Cheng, G., Yang, C., Yao, X., Li, K., and Han, J. (2018). When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs. *IEEE Trans. Geosci. Remote Sens.* 56, 2811–2821. doi: 10.1109/tgrs.2017.2783902

Dong, C., Loy, C. C., He, K., and Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 295–307.

Hou, Y., Li, Z., Wang, P., and Zhang, W. (2018). Skeleton optical spectra-based action recognition using convolutional neural networks. *IEEE Trans. Circuits Syst. Video Technol.* 28, 807–811. doi: 10.1109/tcsvt.2016.2628339

Huang, H., Yang, J., Huang, H., Song, Y., and Gui, G. (2018). Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system. *IEEE Trans. Veh. Technol.* 67, 8549–8560. doi: 10.1109/tvt.2018.2851783

Jiao, R., Huang, X., Ma, X., Han, L., and Tian, W. (2018). A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. *IEEE Access* 6, 17851–17858. doi: 10.1109/access.2018.2818108

Kosiorek, A. R., Sabour, S., Teh, Y. W., and Hinton, G. E. (2019). Stacked capsule autoencoders. *arXiv* [Preprint]. 1906.06818

Li, H., and Wu, X. (2019). DenseFuse: a fusion approach to infrared and visible images. *IEEE Trans. Image Process.* 28, 2614–2623. doi: 10.1109/tip.2018.2887342

Liang, R. Z., Shi, L., Wang, H., Meng, J., Wang, J. J. Y., Sun, Q., et al. (2016). "Optimizing top precision performance measure of content-based image retrieval by learning similarity function," in *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR)* (Cancun: IEEE), 2954–2958.

Liu, D., Wang, Z., Fan, Y., Liu, X., Wang, Z., Chang, S., et al. (2018). Learning temporal dynamics for video super-resolution: a deep learning approach. *IEEE Trans. Image Process.* 27, 3432–3445. doi: 10.1109/tip.2018.2820807

Liu, Y., Chen, X., Peng, H., and Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. *Inform. Fusion* 36, 191–207. doi: 10.1016/j.inffus.2016.12.001

Liu, Y., Chen, X., Ward, R. K., and Wang, Z. J. (2019). Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Process. Lett.* 26, 485–489. doi: 10.1109/lsp.2019.2895749

Ma, J., Yu, W., Liang, P., Li, C., and Jiang, J. (2019). FusionGAN: a generative adversarial network for infrared and visible image fusion. *Inform. Fusion* 48, 11–26. doi: 10.1016/j.inffus.2018.09.004

Ma, K., Zeng, K., and Wang, Z. (2015). Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* 24, 3345–3356. doi: 10.1109/tip.2015.2442920

Ma, S., Chen, M., Wu, J., Wang, Y., Jia, B., and Jiang, Y. (2018). High-voltage circuit breaker fault diagnosis using a hybrid feature transformation approach based on random forest and stacked auto-encoder. *IEEE Trans. Ind. Electron.* 66, 9777–9788. doi: 10.1109/tie.2018.2879308

Pan, Z. W., and Shen, H. L. (2019). Multispectral image super-resolution via rgb image fusion and radiometric calibration. *IEEE Trans. Image Process.* 28, 1783–1797. doi: 10.1109/tip.2018.2881911

Piella, G., and Heijmans, H. (2003). "A new quality metric for image fusion," in *Processings of the IEEE International Conference on Image Processing*, Vol. 3 (Barcelona: IEEE), 173–176.

Schlemper, J., Caballero, J., Hajnal, J. V., Price, A., and Rueckert, D. (2018). A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging* 37, 491–503. doi: 10.1109/tmi.2017.2760978

Shao, Z., and Cai, J. (2018). Remote sensing image fusion with deep convolutional neural network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11, 1656–1669.

Sun, G., Huang, H., Zhang, A., Li, F., Zhao, H., and Fu, H. (2019). Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images. *Remote Sens.* 11:227. doi: 10.3390/rs11030227

Takaishi, D., Kawamoto, Y., Nishiyama, H., Kato, N., Ono, F., and Miura, R. (2018). Virtual cell based resource allocation for efficient frequency utilization in unmanned aircraft systems. *IEEE Trans. Veh. Technol.* 67, 3495–3504. doi: 10.1109/tvt.2017.2776240

Wang, Z., and Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Process. Lett.* 9, 81–84. doi: 10.1109/97.995823

Ye, D., Fuh, J. Y. H., Zhang, Y., Hong, G. S., and Zhu, K. (2018). In situ monitoring of selective laser melting using plume and spatter signatures by deep belief networks. *ISA Trans.* 81, 96–104. doi: 10.1016/j.isatra.2018.07.021

Yin, H., Li, S., and Fang, L. (2013). Simultaneous image fusion and super-resolution using sparse representation. *Inform. Fusion* 14, 229–240. doi: 10.1016/j.inffus.2012.01.008

Yizhang, J., Kaifa, Z., Kaijian, X., Jing, X., Leyuan, Z., Yang, D., et al. (2019). A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation. *J. Med. Syst.* 43:118.

Yizhang, J., Xiaoqing, G., Dongrui, W., Wenlong, H., Jing, X., Shi, Q., et al. (2021). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 40–52. doi: 10.1109/TCBB.2019.2963873

Zhang, Q., Liu, Y., Blum, R. S., Han, J., and Tao, D. (2018). Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: a review. *Inform. Fusion* 40, 57–75. doi: 10.1016/j.inffus.2017.05.006

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 June 2021
doi: 10.3389/fnins.2021.634926

Characterizing Brain Tumor Regions Using Texture Analysis in Magnetic Resonance Imaging

Yun Yu[1,2†], Xi Wu[1†], Jiu Chen[2], Gong Cheng[3], Xin Zhang[1,4], Cheng Wan[1], Jie Hu[1], Shumei Miao[1,4], Yuechuchu Yin[1,4], Zhongmin Wang[1,4], Tao Shan[1,4], Shenqi Jing[1,4], Wenming Wang[1,4], Jianjun Guo[1,4], Xinhua Hu[5] and Yun Liu[1,4]**

[1] School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China, [2] Institute of Brain Functional Imaging, Nanjing Medical University, Nanjing, China, [3] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China, [4] Institute of Medical Informatics and Management, Nanjing Medical University, Nanjing, China, [5] Department of Neurosurgery, The Affiliated Brain Hospital of Nanjing Medical University, Nanjing, China

OPEN ACCESS

Edited by:
Kaijian Xia,
Changshu No.1 People's Hospital, China

Reviewed by:
Ling Tao,
Nanjing University of Aeronautics and Astronautics, China
Tongguang Ni,
Changzhou University, China

***Correspondence:**
Yun Liu
liuyun@njmu.edu.cn
Xinhua Hu
huxinhua1978@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 29 November 2020
Accepted: 06 April 2021
Published: 03 June 2021

Citation:
Yu Y, Wu X, Chen J, Cheng G, Zhang X, Wan C, Hu J, Miao S, Yin Y, Wang Z, Shan T, Jing S, Wang W, Guo J, Hu X and Liu Y (2021) Characterizing Brain Tumor Regions Using Texture Analysis in Magnetic Resonance Imaging. Front. Neurosci. 15:634926. doi: 10.3389/fnins.2021.634926

Purpose: To extract texture features from magnetic resonance imaging (MRI) scans of patients with brain tumors and use them to train a classification model for supporting an early diagnosis.

Methods: Two groups of regions (control and tumor) were selected from MRI scans of 40 patients with meningioma or glioma. These regions were analyzed to obtain texture features. Statistical analysis was conducted using SPSS (version 20.0), including the Shapiro–Wilk test and Wilcoxon signed-rank test, which were used to test significant differences in each feature between the tumor and healthy regions. T-distributed stochastic neighbor embedding (t-SNE) was used to visualize the data distribution so as to avoid tumor selection bias. The Gini impurity index in random forests (RFs) was used to select the top five out of all features. Based on the five features, three classification models were built respectively with three machine learning classifiers: RF, support vector machine (SVM), and back propagation (BP) neural network.

Results: Sixteen of the 25 features were significantly different between the tumor and healthy areas. Through the Gini impurity index in RFs, standard deviation, first-order moment, variance, third-order absolute moment, and third-order central moment were selected to build the classification model. The classification model trained using the SVM classifier achieved the best performance, with sensitivity, specificity, and area under the curve of 94.04%, 92.3%, and 0.932, respectively.

Conclusion: Texture analysis with an SVM classifier can help differentiate between brain tumor and healthy areas with high speed and accuracy, which would facilitate its clinical application.

Keywords: texture analysis, MRI, SVM, brain tumor, t-SNE

INTRODUCTION

Brain cancer remains a diagnostic challenge for clinicians and radiologists because malignant brain tumor cells can invade into the neighboring cells in the brain and spinal cord with fuzzy borders and have a high progression rate (Wild, 2014; Vargo, 2017; Tandel et al., 2019). Treatment of advanced brain tumors is difficult; therefore, early diagnosis is of great importance in clinical settings. The approaches currently employed for the diagnosis of brain tumors include both invasive and noninvasive methods. Although the invasive diagnostic method—biopsy—is viewed as the golden standard for the diagnosis of brain tumors, noninvasive diagnostic methods including magnetic resonance imaging (MRI) are safer and more widely used (Zhao and Jia, 2016). Determination of the accurate location and segmentation of the brain tumor on MRI scans are essential for treatment planning (Mahaley et al., 1989). Several studies have found MRI features capable of differentiating between the tumor and healthy regions (May et al., 1991; Drape et al., 1992; Mullen and Huang, 2017). However, in most cases, the diagnostic accuracy only depends on the proficiency of the medical practitioner reading the MRI scan (Hayward et al., 2008). Many complex patterns, also called image textures, remain imperceptible to the naked eye. Texture analysis is a practical approach for image pattern recognition by extracting objective information through the analysis of the spatial distribution of intensity variations on images (Haralick and Shanmugam, 1973; Haralick, 1979). Furthermore, several studies have confirmed the efficiency of texture analysis (Bayanati et al., 2015; Hodgdon et al., 2015; Skogen et al., 2016).

To increase the diagnostic precision and efficiency, many computer-assisted methods have been developed and introduced, including machine learning (ML) and deep learning (DL) (Zhao and Jia, 2016; Boissoneault et al., 2017; Salvador et al., 2017). Texture analysis combined with ML methods has been widely used to evaluate medical images and yielded promising results (Fetit et al., 2015; Li et al., 2016; Bisdas et al., 2018). However, to the best of our knowledge, there are a few reports on the use of t-distributed stochastic neighbor embedding (t-SNE), which is a new dimensionality reduction and visualization technique to foresee data for preventing problems such as incorrect marking of images and that can help increase the accuracy of the classification.

We hypothesized that some texture features acquired from MRI scans would serve as classification features and markedly improve classification efficiency. To test our hypothesis, the Gini impurity index in the random forests (RFs) was applied to select features, which were then used to develop classification models. Finally, the performance of the features and the models in confirming our hypothesis was assessed.

SUBJECTS AND METHODS

Subjects

The data used were collected from the Affiliated Nanjing Brain Hospital of Nanjing Medical University. Patients in whom meningioma or glioma was histopathologically confirmed between January 2014 and December 2014 were selected. In all, 40 patients (average age: 51.10 years) comprising 22 men (average age: 52.36 years) and 18 women (average age: 47.33 years) were included. The exclusion criteria were as follows: (1) presence of other organic mental disorders and nervous system diseases and (2) a history of major physical illnesses. All of the patients met the above criteria. The study was approved by the medical ethics committee of Nanjing Medical University. All patients provided signed written informed consent.

MRI Acquisition

All images were acquired using a 3T Siemens MRI system. The patients were instructed to relax, keep their eyes closed, stay awake, and remain still. Patient compliance was confirmed after scanning was completed. The images were recorded axially for 6 min by using an echo-planar imaging sequence with the following parameters: TR = 1900 ms, TE = 2.49 ms, slice thickness = 1 mm, flip angle = 90°, and matrix size = 256 × 256. All patients underwent MRI without reporting discomfort during or after the procedure.

Classification Based on 25 Texture Features

Preparation Before Classification

For the experimental preparation, the raw sample image format was changed from DICOM to JPG. In the texture analysis, the tumor region in the coronal MRI image was selected as the experimental group, and the symmetrical healthy region on the other side of the brain was selected as the control group. There were 40 tumor regions in the experimental group and 40 healthy regions in the control group. In each group, 25 texture features (belonging to three categories) were calculated, as shown in **Table 1**.

The 25 texture features were recorded as mean ± SD. Statistical analysis was performed using SPSS (version 20.0), including the Shapiro–Wilk test and Wilcoxon signed-rank test, which was used for testing significant differences in each feature between tumors and healthy areas. Meanwhile, an RF model was employed to predict whether each sample was a tumor or a healthy area and for importance rankings of 25 texture features according to the Gini impurity index in the RF (Menze et al., 2009; Liu et al., 2018). All texture features were selected as predictors to compare the experimental results from the Wilcoxon signed-rank test and RF prediction. In addition,

TABLE 1 | The selected texture features in the three categories.

Category (-based parameters)	Texture features
Histogram	(1) First-order moment; (2) second-order moment; (3) third-order moment; (4) fourth-order moment; (5) the central moment of the four features; (6) the absolute moment of the four features
Run-length matrix	(1) Long run emphasis; (2) short run emphasis; (3) gray level nonuniformity; (4) total run-length percentage
Co-occurrence matrix	(1) Energy; (2) inertia moment; (3) correlation; (4) entropy; (5) mean; (6) variance; (7) standard deviation; (8) homogeneity; (9) dissimilarity

TABLE 2 | Wilcoxon signed-rank test results (histogram).

No.	Texture features	Tumor region	Health region	*P*-value
1	First-order moment	10.661 ± 3.414	3.664 ± 5.111	0.000*
2	First-order central moment	2.327 ± 0.280	1.071 ± 0.934	0.000*
3	First-order absolute moment	2.327 ± 1.280	1. 071 ± 0.934	0.000*
4	Second-order moment	159.912 ± 51.205	54.958 ± 76.657	0.000*
5	Second-order central moment	19,442.302 ± 13,715.926	6,490.148 ± 13,691.211	0.000*
6	Second-order absolute moment	19,442.302 ± 13,715.926	6,490.148 ± 13,691.211	0.000*
7	Third-order moment	2,398.676 ± 768.069	824.368 ± 1,149.855	0.000*
8	Third-order central moment	−14,586,918,376.546 ± 12,193,828,145.790	−5,165,830,452.742 ± 11,313,615,571.915	0.000*
9	Third-order absolute moment	14,586,918,376.546 ± 12,193,828,145.790	5,165,830,452.768 ± 11,313,615,571.903	0.000*
10	Fourth-order moment	35,980.138 ± 11,521.031	12,365.514 ± 17,247.826	0.000*
11	Fourth-order central moment	2,201,613,996,782,910,720.000 ± 2,158,254,619,227,325,440.000	827,213,133,922,581,760.000 ± 1,882,686,712,587,172,610.000	0.000*
12	Fourth-order absolute moment	2,201,613,996,782,910,720.000 ± 2,158,254,619,227,325,440.000	827,213,133,922,581,760.000 ± 1,882,686,712,587,172,610.000	0.000*

**means that this data shows significance.*

t-SNE, a new dimension reduction and visualization technique for high-dimensionality data, was performed in the exploratory analysis (Li et al., 2017). It was applied to all 40 pairs of samples with 25 features to delete the samples that would apparently have a negative effect on the latter classification.

Classification

The samples were randomly divided into training (70%) and test sets (30%). This was iterated five times to provide five unique training and testing groups. The training set was used to generate classification models with three different classifiers: RF, BP, and SVM.

The RF is fast, is flexible, and has become a standard tool in biomedical informatics. Each classifier in the ensemble is a decision tree classifier and is generated using random selection of attributes at each node to determine the split.

TABLE 3 | Wilcoxon signed-rank test results (run-length matrix).

No.	Texture features	Tumor region	Healthy region	*P*-value
1	Long run emphasis	628.833 ± 512.533	782.519 ± 639.539	0.188
2	Short run emphasis	0.228 ± 0.076	0.234 ± 0.100	0.582
3	Total run-length percentage	0.084 ± 0.0385	0.078 ± 0.046	0.476
4	Gray level nonuniformity	207.211 ± 129.511	154.085 ± 86.213	0.011*

**means that this data shows significance.*

TABLE 4 | Wilcoxon signed-rank test results (co-occurrence matrix).

No.	Texture features	Tumor region	Healthy region	*P*-value
1	Energy	0.637 ± 0.198	0.812 ± 0.161	0.000*
2	Entropy	0.663 ± 0.328	0.404 ± 0.288	0.002*
3	Inertia moment	15.809 ± 10.322	13.5043 ± 12.0488	0.313
4	Correlation	0.105 ± 0.276	0.093 ± 0.150	0.026*
5	Homogeneity	0.939 ± 0.040	0.948 ± 0.047	0.313
6	Dissimilarity	0.988 ± 0.645	0.844 ± 0.753	0.313
7	Mean	12.444 ± 3.645	4.876 ± 5.465	0.000*
8	Variance	152.277 ± 47.371	53.902 ± 71.028	0.000*
9	Standard deviation	15.004 ± 2.084	8.575 ± 4.677	0.000*

**means that this data shows significance.*

TABLE 5 | Gini impurity index in the RF.

No.	Texture features	Mean decrease Gini	Rank
1	First-order moment	1.7897453	7
2	First-order central moment	0.8495841	16
3	First-order absolute moment	0.7873654	19
4	Second-order moment	1.8026458	6
5	Second-order central moment	1.5147000	10
6	Second-order absolute moment	1.3757714	13
7	Third-order moment	1.6430335	9
8	Third-order central moment	1.8037972	5
9	Third-order absolute moment	1.8316361	4
10	Fourth-order moment	2.0892720	2
11	Fourth-order central moment	1.0609800	14
12	Fourth-order absolute moment	1.4124922	12
13	Energy	0.4763909	25
14	Entropy	0.5165798	23
15	Inertia moment	0.7618972	20
16	Correlation	0.7355067	21
17	Homogeneity	0.8385106	17
18	Dissimilarity	0.7343357	22
19	Mean	1.7526929	8
20	Variance	1.9482926	3
21	Standard deviation	3.1126458	1
22	Long-run emphasis	0.8664384	15
23	Short-run emphasis	1.4297721	11
24	Total run-length percentage	0.8314745	18
25	Gray-level nonuniformity	0.4918337	24

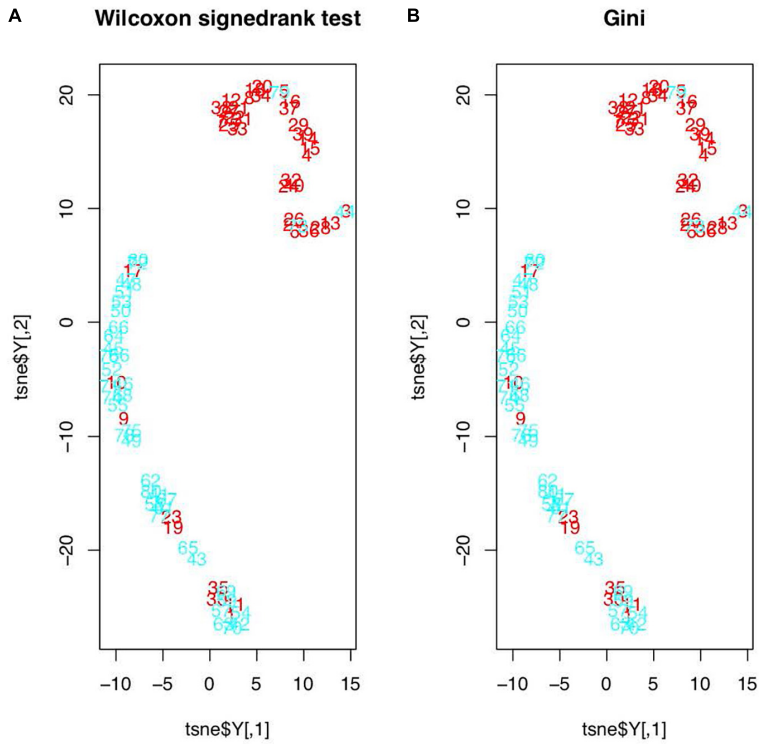


FIGURE 1 | Feature distributions after t-SNE: **(A)** features based on the Wilcoxon signed-rank test; **(B)** features based on the RF's importance rankings (red figures represent the healthy regions and blue figures represent the tumor regions).

During classification, each tree votes, and the most popular class is returned.

The BP iteratively processes a set of training tuples and compares the network's prediction with the actual known target value. For each training tuple, the weights are modified to minimize the mean squared error between the network's prediction and the actual target value. Modifications are made in the backwards direction. The process will reach the terminating condition when the error is very small.

The SVM is a classification method for both linear and nonlinear data. It uses nonlinear mapping to transform the original training data into a higher dimension. With the new dimension, it searches for the linear optimal separating hyperplane. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane. SVM finds this hyperplane using support vectors and margins.

Four indexes were used to evaluate each model, including the area under the curve (AUC), error rate, sensitivity, and specificity.

TABLE 6 | Three classifiers evaluation.

Classifiers	AUC	Error rate (%)	Sensitivity (%)	Specificity (%)
RF	0.856	14.1	82.8	88.3
SVM	0.932	6.9	94.04	92.3
BP	0.884	11.4	87.2	89.6

Moreover, the receiver operating characteristic (ROC) curve was constructed for each model.

RESULTS

Texture Feature Analysis

Using the Wilcoxon signed-rank test, 25 texture features could reveal regions with higher and lower texture values when comparing the experimental (tumor region) and control groups (health region), as shown in **Tables 2–4**. We obtained the importance rankings of the 25 texture features according to the Gini impurity index in the RF with a training set (80%). The top five features were standard deviation, first-order moment, variance, third-order absolute moment, and third-order central moment, as shown in **Table 5**.

The t-SNE test results are shown in **Figure 1**. In **Figure 1A**, the original features were those found in the Wilcoxon signed-rank test (19 features in total), and in **Figure 1B**, the original features were the top five features determined in the RF's importance rankings. However, the data distributions after t-SNE were similar. All samples were evidently divided into two clusters, except 12 samples (1, 9, 10, 11, 17, 19, 23, 30, 35, 44, 73, and 79), which were seemingly distributed mistakenly. In addition, *t*-test was used to examine 40 samples to determine whether their features were relatively different between the tumor and healthy regions. We found that the mean *P*-value was

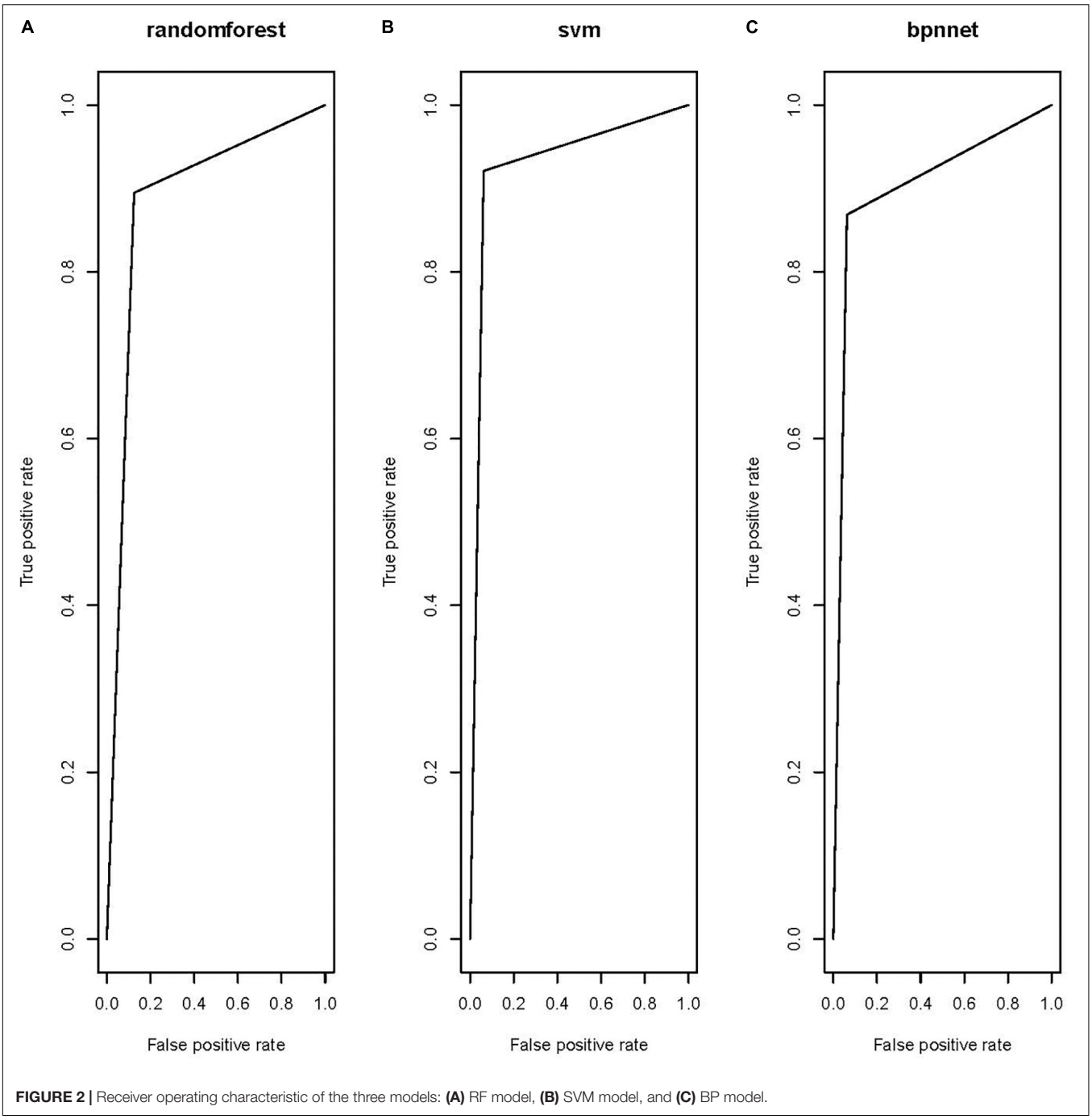


FIGURE 2 | Receiver operating characteristic of the three models: **(A)** RF model, **(B)** SVM model, and **(C)** BP model.

0.2390645. The *P*-values of seven samples—1, 9, 10, 19, 23, 30, and 35—were greater than the mean *P*-value, and these samples were also mistakenly distributed in the t-SNE picture and were deleted.

Classifier Evaluation

On the basis of the results obtained above, we selected the five features (standard deviation, first-order moment, variance, third-order absolute moment, and third-order central moment) identified in the RF to set up classifiers, which helped save calculation time and resources. Three classification models (RF, SVM, and BP) were applied, and five features were used to train each classifier. A detailed summary of the model's performance is presented in **Table 6**.

All three models showed satisfactory AUCs of 0.85–0.95. The RF and the BP shared a similar performance based on the AUC, error rate, sensitivity, and specificity. The model trained by the SVM classifier demonstrated the best performance among the three models, with markedly better AUC, error rate, sensitivity, and specificity, indicating that this model

could correctly classify the tumor and healthy regions. Receiver operating characteristic (ROC) curves were constructed for the three models to compare their performance directly, as shown in **Figure 2**.

DISCUSSION

Some studies have reported the same methods to select features, and the validity of this approach has been proven. Wang et al. (2018) evaluated the importance of spectra lines based on RFs and then used a support vector machine (SVM) classifier to classify the laser-induced plasma spectra (LIBS) of bacteria species. The primary objective of this study was to characterize tumor regions using MRI-based texture analysis. We used texture analysis to compute 25 texture features from MRI images. Using the Wilcoxon signed-rank test, we confirmed that 19 texture features of the total 25 features were different between the healthy and tumor regions. Using the t-SNE technique, the dataset was divided into two clusters, indicating that there is a high possibility to set up a classification model with these 19 features. However, training a model with high-dimensionality data requires a lot of time and space. To facilitate faster and more accurate classification, the importance rankings of the features in the RF were calculated, and the top five features were found to show the same classification effectiveness as the 19 features selected before.

The images for the t-SNE test results showed some seemingly noisy dots. Considering the possibility that all mistakenly distributed samples may be deleted incorrectly, the *t*-test was applied to generally examine whether the healthy and tumor regions showed significant differences in the 25 texture features for each sample. To determine the modified number of samples that would be deleted, the mean *P*-value was set as the deletion standard, and seven samples were excluded on the basis of this standard. Since the samples were marked manually and these samples were likely to be marked mistakenly, this was a limitation that has been mentioned in many previous studies.

On the basis of the five features, three class-action models were built by training three ML classifiers, namely, RF, SVM, and BP. The SVM classifier was superior to the RF and BP classifiers, as shown in **Table 6**, since it provided better performance in terms of AUC, error rate, sensitivity, and specificity. These results were shown to be convincing through fivefold confirmation, which was consistent with the findings of previous studies (Zhang et al., 2017). The model in this article was superior to the previous models since it depended on only five features while showing the same AUC. Since the software that is needed to perform texture analysis and build classification models is readily available, clinicians can easily perform such analyses in clinical settings.

This study had some limitations. First, the dataset was modified, since the model was trained with only 80 samples. Its robustness needs further examination. Second, some degree of selection bias may exist. Different categories of brain tumors have different texture features. Some unique features were excluded, which may have influenced the results of our analysis. Third, a manual approach was adopted to segment tumors in this study. Although manual segmentation generally works better than automatic methods, segmentation errors could still exist. Some noise dots may have been mistakenly marked manually, negatively influencing the formation of our model.

CONCLUSION

In conclusion, we hypothesized that a few of the textures acquired from the MRI images could serve the role of classification features, thereby significantly improving the classification efficiency. The Gini impurity index in the RF was applied to select features. On the basis of the five features, three class-action models were built by training three ML classifiers, including RF, SVM, and BP. The classifier model in this article was superior to the previous models, since it depended on only five features. On the basis of our initial findings, tumor regions characterized on the basis of MRI-based texture analysis may have clinical usefulness in differentiating brain tumors.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data cannot be used out of the hospital. Requests to access the datasets should be directed to med.info@njmu.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by grants from the National Key Research & Development Plan of the Ministry of Science and Technology of the People's Republic of China (Grant nos. 2018YFC1314900 and 2018YFC1314901), 2019 provincial special guide fund project for the Development of Modern Service Industry [2019 (783)], the 2018 project of Jiangsu Province Department of Industry and Information Technology (Grant no. 2018419), the 2018 project of Jiangsu Provincial University Philosophy and Social Science Research Fund (2018SJA0282), and the 2016 projects of Nanjing Science Bureau (Grant no. 201608003). YL was the guarantor of this article.

REFERENCES

Bayanati, H., Thornhill, R. E., Souza, C. A., Sethi-Virmani, V., Gupta, A., Maziak, D., et al. (2015). Quantitative CT texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? *Eur. Radiol.* 25, 480–487. doi: 10.1007/s00330-014-3420-6

Bisdas, S., Shen, H., Thust, S., Katsaros, V., Stranjalis, G., Boskos, C., et al. (2018). Texture analysis- and support vector machine-assisted diffusional kurtosis imaging may allow in vivo gliomas grading and IDH-mutation status prediction: a preliminary study. *Sci. Rep.* 8:6108. doi: 10.1038/s41598-018-24438-4

Boissoneault, J., Sevel, L., Letzen, J., Robinson, M., and Staud, R. (2017). Biomarkers for musculoskeletal pain conditions: use of brain imaging and machine learning. *Curr. Rheumatol. Rep.* 19:5. doi: 10.1007/s11926-017-0629-9

Drape, J. L., Krause, D., and Tongio, J. (1992). MRI of Aggressive Meningiomas. *J. Neuroradiol.* 19, 49–62.

Fetit, A. E., Novak, J., Rodriguez, D., Auer, D. P., Clark, C. A., Grundy, R. G., et al. (2015). 3D texture analysis of heterogeneous MRI data for diagnostic classification of childhood brain tumours. *Stud. Health Technol. Inform.* 213, 19–22.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proc. IEEE* 67, 786–804. doi: 10.1109/proc.1979.11328

Haralick, R. M., and Shanmugam, K. (1973). Textural features for image classification. *IEEE Trans. Syst. Man. Cybern* 3, 610–621. doi: 10.1109/TSMC.1973.4309314

Hayward, R. M., Patronas, N., Baker, E. H., Vézina, G., Albert, P. S., and Warren, K. E. (2008). Inter-observer variability in the measurement of diffuse intrinsic pontine gliomas. *J. Neuro. Oncol.* 90, 57–61. doi: 10.1007/s11060-008-9631-4

Hodgdon, T., McInnes, M. D., Schieda, N., Flood, T. A., Lamb, L., and Thornhill, R. E. (2015). Can quantitative CT texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced CT images? *Radiology* 276, 787–796. doi: 10.1148/radiol.2015142215

Li, W., Cerise, J. E., Yang, Y., and Han, H. (2017). Application of T-SNE to human genetic data. *J. Bioinform. Comput. Biol.* 15:1750017. doi: 10.1142/S0219720017500172

Li, Y., Li, M. M., Zhang, Y., Cheng, J. L., Shang, Z. G., and Bu, C. X. (2016). [Utility of texture analysis of magnetic resonance imaging in differential diagnosis of common pediatric cerebellar tumors in children]. *Zhonghua Yi Xue Za Zhi* 96, 1853–1855.

Liu, Y., Zhang, Y., Liu, D., Tan, X., Tang, X., Zhang, F., et al. (2018). Prediction of ESRD in IgA nephropathy patients from an asian cohort: a random forest model. *Kidney Blood Press. Res.* 43, 1852–1864. doi: 10.1159/000495818

Mahaley, M. S. Jr., Mettlin, C., Natarajan, N., Laws, E. R. Jr., and Peace, B. B. (1989). National survey of patterns of care for brain-tumor patients. *J. Neurosurg.* 71, 826–836. doi: 10.3171/jns.1989.71.6.0826

May, P. L., Blaser, S. I., Hoffman, H. J., Humphreys, R. P., and Harwood-Nash, D. C. (1991). Benign intrinsic tectal 'tumors' in children. *J. Neurosurg.* 74, 867–871. doi: 10.3171/jns.1991.74.6.0867

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., et al. (2009). A Comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 10:213. doi: 10.1186/1471-2105-10-213

Mullen, K. M., and Huang, R. Y. (2017). An update on the approach to the imaging of brain tumors. *Curr. Neurol. Neurosci. Rep.* 17:53. doi: 10.1007/s11910-017-0760-z

Salvador, R., Radua, J., Canales-Rodríguez, E. J., Solanes, A., Sarró, S., Goikolea, J. M., et al. (2017). Evaluation of machine learning algorithms and structural features for optimal mri-based diagnostic prediction in psychosis. *PLoS One* 12:e0175683. doi: 10.1371/journal.pone.0175683

Skogen, K., Schulz, A., Dormagen, J. B., Ganeshan, B., Helseth, E., and Server, A. (2016). Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. *Eur. J. Radiol.* 85, 824–829. doi: 10.1016/j.ejrad.2016.01.013

Tandel, G. S., Biswas, M., Kakde, O. G., Tiwari, A., Suri, H. S., Turk, M., et al. (2019). A review on a deep learning perspective in brain cancer classification. *Cancers* 11:111. doi: 10.3390/cancers11010111

Vargo, M. M. (2017). Brain tumors and metastases. *Phys. Med. Rehabil. Clin. North Am.* 28, 115–141. doi: 10.1016/j.pmr.2016.08.005

Wang, Q., Teng, G., Qiao, X., Zhao, Y., Kong, J., Dong, L., et al. (2018). Importance evaluation of spectral lines in laser-induced breakdown spectroscopy for classification of pathogenic bacteria. *Biomed. Opt. Express.* 9, 5837–5850. doi: 10.1364/BOE.9.005837

Wild, C. P. (2014). International agency for research on cancer. *Encyclopedia Toxicol.* 133, 1067–1069. doi: 10.1016/b978-0-12-386454-3.00402-4

Zhang, Y., Cheng, J., and Li, Y. (2017). The utility of texture analysis of magnetic resonance imaging in defining the boundaries of high grade gliomas. *J. Clin. Radiol.* 36, 315–318.

Zhao, L., and Jia, K. (2016). Multiscale CNNs for brain tumor segmentation and diagnosis. *Comput. Math. Methods Med.* 2016:8356294. doi: 10.1155/2016/8356294

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.

Copyright © 2021 Yu, Wu, Chen, Cheng, Zhang, Wan, Hu, Miao, Yin, Wang, Shan, Jing, Wang, Guo, Hu 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 in Neuroscience

ORIGINAL RESEARCH
published: 09 June 2021
doi: 10.3389/fnins.2021.634557

Disrupted Asymmetry of Inter- and Intra-Hemispheric Functional Connectivity at Rest in Medication-Free Obsessive-Compulsive Disorder

Cuicui Jia[1†], Yangpan Ou[2†], Yunhui Chen[1†], Jidong Ma[3], Chuang Zhan[3], Dan Lv[1], Ru Yang[4], Tinghuizi Shang[1], Lei Sun[1], Yuhua Wang[1], Guangfeng Zhang[5], Zhenghai Sun[1], Wei Wang[6], Xiaoping Wang[2], Wenbin Guo[2] and Ping Li[1]**

[1] Department of Psychiatry, Qiqihar Medical University, Qiqihar, China, [2] National Clinical Research Center for Mental Disorders, and Department of Psychiatry, The Second Xiangya Hospital of Central South University, Changsha, China, [3] Department of Psychiatry, Baiyupao Psychiatric Hospital of Harbin, Harbin, China, [4] Department of Radiology, The Second Xiangya Hospital of Central South University, Changsha, China, [5] Department of Radiology, The Third Affiliated Hospital of Qiqihar Medical University, Qiqihar, China, [6] Department of Library, Qiqihar Medical University, Qiqihar, China

Disrupted functional asymmetry of cerebral hemispheres may be altered in patients with obsessive-compulsive disorder (OCD). However, little is known about whether anomalous brain asymmetries originate from inter- and/or intra-hemispheric functional connectivity (FC) at rest in OCD. In this study, resting-state functional magnetic resonance imaging was applied to 40 medication-free patients with OCD and 38 gender-, age-, and education-matched healthy controls (HCs). Data were analyzed using the parameter of asymmetry (PAS) and support vector machine methods. Patients with OCD showed significantly increased PAS in the left posterior cingulate cortex, left precentral gyrus/postcentral gyrus, and right inferior occipital gyrus and decreased PAS in the left dorsolateral prefrontal cortex (DLPFC), bilateral middle cingulate cortex (MCC), left inferior parietal lobule, and left cerebellum Crus I. A negative correlation was found between decreased PAS in the left DLPFC and Yale–Brown Obsessive-compulsive Scale compulsive behavior scores in the patients. Furthermore, decreased PAS in the bilateral MCC could be used to distinguish OCD from HCs with a sensitivity of 87.50%, an accuracy of 88.46%, and a specificity of 89.47%. These results highlighted the contribution of disrupted asymmetry of intra-hemispheric FC within and outside the cortico-striato-thalamocortical circuits at rest in the pathophysiology of OCD, and reduced intra-hemispheric FC in the bilateral MCC may serve as a potential biomarker to classify individuals with OCD from HCs.

Keywords: obsessive-compulsive disorder, functional magnetic resonance imaging, resting-state, parameter of asymmetry, support vector machine

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Jinping Xu, Chinese Academy of Sciences (CAS), China
Hongru Zhao, Soochow University, China

***Correspondence:**
Ping Li
lipingchxyy@163.com
Wenbin Guo
guowenbin76@csu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *28 November 2020*
Accepted: *07 May 2021*
Published: *09 June 2021*

Citation:
Jia C, Ou Y, Chen Y, Ma J, Zhan C, Lv D, Yang R, Shang T, Sun L, Wang Y, Zhang G, Sun Z, Wang W, Wang X, Guo W and Li P (2021) Disrupted Asymmetry of Inter- and Intra-Hemispheric Functional Connectivity at Rest in Medication-Free Obsessive-Compulsive Disorder. Front. Neurosci. 15:634557. doi: 10.3389/fnins.2021.634557

INTRODUCTION

Asymmetry is considered when one side of the brain is structurally or functionally different from the other side (Zach et al., 2016). Structural and functional asymmetry of cerebral hemispheres exists in the healthy population (Crow et al., 1989), which may be due to hereditary, developmental, evolutionary, pathological, and epigenetic factors (Samara and Tsangaris, 2011). Healthy individuals can benefit from normal asymmetries. However, the asymmetries are disrupted in some mental disorders, such as schizophrenia (Federspiel et al., 2006), attention-deficit/hyperactivity disorder (Shaw et al., 2009), somatization disorder (Su et al., 2020), and autism (Herbert et al., 2005). Patients with obsessive-compulsive disorder (OCD) also displayed anomalous brain asymmetries (Boedhoe et al., 2017). Moreover, patients with OCD and their siblings exhibited leftward asymmetry of cortical thickness in the anterior cingulate cortex, and this abnormal asymmetry was positively correlated with compulsive subscale scores (Peng et al., 2015). Pediatric and adult patients with OCD both showed significantly altered brain volume asymmetries (Wang et al., 2020). For example, pediatric OCD displayed brain volume asymmetry in the pallidum and thalamus (Kong et al., 2020). In addition, reduced interhemispheric functional connectivity (FC) within (i.e., orbital frontal cortex and thalamus) and outside (i.e., inferior occipital gyrus and precentral gyrus/postcentral gyrus) the cortico-striato-thalamocortical (CSTC) circuits in OCD has been discovered by the authors' group (Jia et al., 2020). Taken together, structural and functional asymmetry of cerebral hemispheres exists in patients with OCD.

Given that the corpus callosum connects functional interaction and cooperation between two hemispheres (Schüz and Preissl, 1996; Doron and Gazzaniga, 2008), each individual brain region has FCs related to inter- and intra-hemispheric connections, thus contributing to the information communication of cognition and behavior (Zhu et al., 2019). Despite significant advances in OCD research (Jia et al., 2020), whether the asymmetric functional disruption of OCD resulting from inter- and/or intra-hemispheric FCs and the extent of these FCs contribute to abnormal functional asymmetry in OCD remain unclear.

We have used the voxel-mirrored homotopic connectivity (VMHC) method to investigate the interhemispheric coordination at rest in OCD in a previous study (Jia et al., 2020), whereas the quantitative parameter of asymmetry (PAS) was used to calculate an abnormal voxel and other voxels from inter- and/or intra-hemispheric asymmetries (Zhu et al., 2018) in this study. Both the PAS and VMHC methods involve interhemispheric FC between the two hemispheres of the brain. However, they are used to assess different aspects of interhemispheric FC. VMHC is applied to assess FC between a given voxel in one hemisphere and its corresponding voxel in the other hemisphere, whereas interhemispheric FC to quantify PAS is utilized to measure a given voxel in one hemisphere and all voxels in the other hemisphere. PAS is a voxel-wise approach that is unaffected by preselected regions of interest (ROIs) and thresholds, thus minimizing the potential confounding effects of structural asymmetries and selection bias through ROIs (Zhu et al., 2018). It has been widely used in exploring inter- and/or intra-hemispheric asymmetries in patients with schizophrenia and their unaffected siblings, individuals at ultra-high risk for psychosis, major depressive disorder (Ding et al., 2021), and somatization disorder (Zhu et al., 2018, 2019; Su et al., 2020).

In the current research, functional asymmetries of the whole-brain regions at rest in medication-free patients with OCD were investigated using the PAS approach to address the issue whether the asymmetric disruption of OCD results from inter- and/or intra-hemispheric FCs. Based on previous studies, patients with OCD were hypothesized to have significantly abnormal inter- and/or intra-hemispheric FCs within and outside the CSTC circuits. In addition, disrupted asymmetries may be related to clinical characteristics and can be used as potential neurobiological markers to distinguish OCD from healthy controls (HCs).

EXPERIMENTAL PROCEDURES

Participants

Forty subjects (27 males and 13 females) with OCD were recruited from the Fourth Affiliated Hospital of Qiqihar Medical University and Qiqihar Mental Health Center, China. The diagnosis was based on the Structured Clinical Interview for DMS-IV (SCID), patient version (First et al., 1996). The Yale–Brown Obsessive-Compulsive Scale (Y-BOCS), Hamilton Anxiety Rating Scale (HAMA), and 17-item Hamilton Depression Rating Scale (HAMD) were used to assess clinical symptoms of OCD. A total of 22 patients had a history of antipsychotic, anti-obsessive medication, or antidepressant use, whereas 18 patients were drug naive. None of the patients took any psychotropic drugs for at least 4 weeks before brain-image acquisition. Thirty-eight HCs (25 males and 13 females) that matched with the patients in gender, education, and age were enrolled from the local community by using the SCID, non-patient version (First et al., 2001). All individuals were right-handed, were Han Chinese, and had the same inclusion criteria as follows: (1) 16–50 years old, (2) no serious physical disease or neurological or psychiatric illness, (3) no drug or alcohol dependence, (4) no pregnancy, (5) no contraindication for an MRI scan, and (6) no movement distance of more than 2 mm or rotation angle of more than 2°. HCs who had a first-degree relative with any mental disorders were excluded.

This study was approved by the Research Ethics Committee of Qiqihar Medical University. The participants were informed of the study procedures, and they signed a written informed consent.

MRI Data Acquisition and Preprocessing

For all participants, resting-state functional magnetic resonance imaging (rs-fMRI) using a 3.0-Tesla GE 750 Signa-HDX scanner (General Electric Healthcare, Waukesha, WI, United States) with a 12-channel standard head coil was conducted at The Third Affiliated Hospital of Qiqihar Medical University. All individuals

were instructed to use foam pads and earplugs to reduce the effect of scanner noise, remain in the supine position, close their eyes, relax, stay awake, and remain motionless (especially the head). The rs-fMRI data were obtained *via* an echo-planar imaging (EPI) sequence: axial slices = 33, repetition time = 2000 ms, echo time = 30 ms, slice thickness = 3.5 mm, inter-slice gap = 0.6 mm, flip angle = 90°, field of view = 200 × 200 mm^2, data matrix = 64 × 64, and 240 volumes (8 min) in total. None of the subjects displayed clinically significant brain structural lesions.

Data were preprocessed with the Data Processing & Analysis for Brain Imaging (DPABI) software (Yan et al., 2016). The first 10 functional volumes were discarded to ensure a steady initial signal and adapt to the environment. Slice time correction and head motion correction were performed for the remaining 230 EPI images. Images were then spatially normalized to a standard Montreal Neurological Institute space and resampled to 3 mm × 3 mm × 3 mm. Afterward, the normalized images were smoothed with a 4-mm full-width half-maximum isotropic Gaussian kernel. Subsequently, the signal was linearly detrended and band-pass filtered (0.01–0.08 Hz) to reduce the covariate effect of high-frequency physiological noises and low-frequency drifts. The signals from the cerebrospinal fluid, white matter, and 24 motion parameters were used as nuisance covariates. The identification of "bad" timepoints was scrubbed using a threshold of 0.2 mm of framewise displacement (FD) and one back and two subsequent neighbors (Power et al., 2012), and the mean FD for each subject was calculated.

Calculation of the PAS Values

Parameter of asymmetry analysis was conducted by using codes. The data were analyzed using the voxel-wise whole-brain analysis to calculate the correlation coefficients between a given voxel and the other voxels from the same hemisphere (intra-hemispheric coefficient) or the opposite hemisphere (interhemispheric coefficient). The intra-hemispheric or interhemispheric FC of this voxel was acquired by taking the mean of intra-hemispheric or interhemispheric coefficients. We used the whole-brain mask to calculate the PAS. Small correlation coefficients between voxels may show the confounding effects in the asymmetric analyses. Therefore, weak correlations were removed with a correlation coefficient threshold ($r > 0.2$) for PAS calculation (Wang et al., 2014; Liu et al., 2015). In addition, only positive correlations were used in the PAS calculation due to ambiguous explanations to negative correlations.

In PAS calculation, only the positive correlations were used, as reported previously (Zhu et al., 2018), and the mean coefficients were transformed to z-values (Buckner et al., 2009). The formula for PAS (Zhu et al., 2018) was as follows:

$$PAS = FC_{inter} - FC_{intra},$$

FC_{inter}, interhemispheric FC and FC_{intra}, intra-hemispheric FC. When the PAS scores were positive, the asymmetry was primarily caused by the interhemispheric FC. When the PAS scores were negative, the asymmetry was mainly caused by the intra-hemispheric FC.

Statistical Analysis

The demographic and clinical data were analyzed using two-sample t-tests and the X^2 test with SPSS version 23.0 (SPSS Inc., Chicago, IL, United States). In each group, the one-sample t-test of PAS was performed in the DPABI software; $p < 0.05$ corrected by Gaussian random field theory (voxel significance: $p < 0.001$, cluster significance: $p < 0.05$) was the significance level. Group differences of PAS were conducted with voxel-wise two-sample t-tests in the DPABI software with $p < 0.05$ corrected by Gaussian random field (GRF). The mean FD values and age were used as covariates to reduce the potential effects of these variables.

Pearson's correlation analyses between the mean PAS scores and clinical variables were conducted in OCD. The Bonferroni-corrected significance level was set at $pc < 0.05$. Twenty-one Pearson's correlations were performed (seven brain regions × three scores of the Y-BOCS). The Bonferroni-corrected significance level was set at $pc < 0.05/21$.

Support Vector Machine Analysis

Support vector machine (SVM) is a well-known and popular supervised machine learning technique with higher accuracy and precision. It effectively defines a set of information as well as the functions of different brain regions to find a maximum boundary delimiter to classify the data, which can differentiate patients from HCs using structural and/or functional neuroimaging data such as rs-fMRI (Jia et al., 2020). The orientation of the hyperplane is as far away from the nearest data point of each class as possible. These closest points are called support vectors (Huang et al., 2018). The exploratory analysis was applied to test whether abnormal PAS could be used to classify patients with OCD and HCs. The process of SVM classification consists of three steps: preparing data for classifier training, training and testing, and performance evaluation (Orrù et al., 2012). The

TABLE 1 | Demographic and clinical characteristics of participants.

	OCD patients (n = 40)	HCs (n = 38)	X^2/t	p
Age (years)	27.28 ± 8.16	27.18 ± 8.33	0.05	0.71
Sex (male/female)	27/13	25/13	0.32	1.00
Education (years)	13.40 ± 2.87	13.74 ± 3.03	−0.50	0.83
Illness duration (months)	66.68 ± 75.54			
Y-BOCS total score	24.90 ± 5.73	1.13 ± 0.88	25.27	<0.01
Y-BOCS obsessive thinking	12.85 ± 4.25	0.37 ± 0.49	17.98	<0.01
Y-BOCS compulsive behavior	12.05 ± 4.62	0.74 ± 0.72	14.92	<0.01
HAMD	8.05 ± 4.40	1.45 ± 0.95	9.04	<0.01
HAMA	10.83 ± 6.55	1.16 ± 1.00	9.00	<0.01
FD	0.04 ± 0.02	0.03 ± 0.01	1.25	0.13
Time points scrubbed out	1.13 ± 2.256	1.00 ± 2.418	0.25	0.95

OCD, obsessive-compulsive disorder; Y-BOCS, Yale–Brown Obsessive-Compulsive Scale; HAMD, 17-item Hamilton Depression Rating Scale; HAMA, Hamilton Anxiety Rating Scale; FD, framewise displacement. Variables of age, education, Y-BOCS total score, subscales score, HAMD score, HAMA score, and FD were tested by two sample t-tests, and the results were indicated by t-values. Categorical data such as gender was tested using a Chi-squared test, and the result was indicated by X^2.

"leave-one-out" cross-validation approach was used to acquire the highest sensitivity and specificity, which involved excluding a single subject from each group and using the remaining subjects to train the classifier (Bu et al., 2019). A detailed description of the SVM can be found in a previous study (Liu et al., 2013).

RESULTS

Demographic and Clinical Data

The demographic and clinical data of participants are displayed in **Table 1**. A total of 40 medication-free patients with OCD and 38 HCs consented to enroll in this research. The two groups did not differ in terms of age ($t = 0.05$, $p = 0.71$), gender ($X^2 = 0.32$, $p = 1.00$), educational level ($t = 0.50$, $p = 0.83$), and FD ($t = 1.25$, $p = 0.13$). Among the clinical characteristics, significant group differences were found in Y-BOCS ($t = 25.27$, $p < 0.01$), HAMD ($t = 9.04$, $p < 0.01$), and HAMA ($t = 9.00$, $p < 0.01$).

Group Differences in PAS

Patients with OCD displayed significantly increased PAS in the left posterior cingulate cortex (PCC), left precentral gyrus/postcentral gyrus, and right inferior occipital gyrus, and decreased PAS in the left dorsolateral prefrontal cortex (DLPFC), bilateral middle cingulate cortex (MCC), left inferior parietal lobule (IPL), and left cerebellum Crus I compared to HCs (as shown in **Figure 1** and **Table 2**). The details of PAS values in the OCD and HC groups are presented in **Supplementary Table 1**. The results of one-sample *t*-tests of PAS in each group are shown in **Supplementary Figure 1**.

Correlation Analysis

A significantly negative correlation was observed between the PAS score in the left DLPFC and Y-BOCS compulsive behavior scores in OCD ($r = -0.318$, $pc = 0.045$, Bonferroni corrected, **Figure 2**).

SVM Analysis

Seven brain regions (1 = left cerebellum Crus I, 2 = bilateral MCC, 3 = left DLPFC, 4 = right inferior occipital gyrus, 5 = left inferior parietal lobule, 6 = left PCC, 7 = left precentral gyrus/postcentral gyrus) with abnormal PAS values were discovered in patients with OCD. SVM analysis was applied to these seven brain regions. The classification accuracies were as follows: 1 = 75.64%

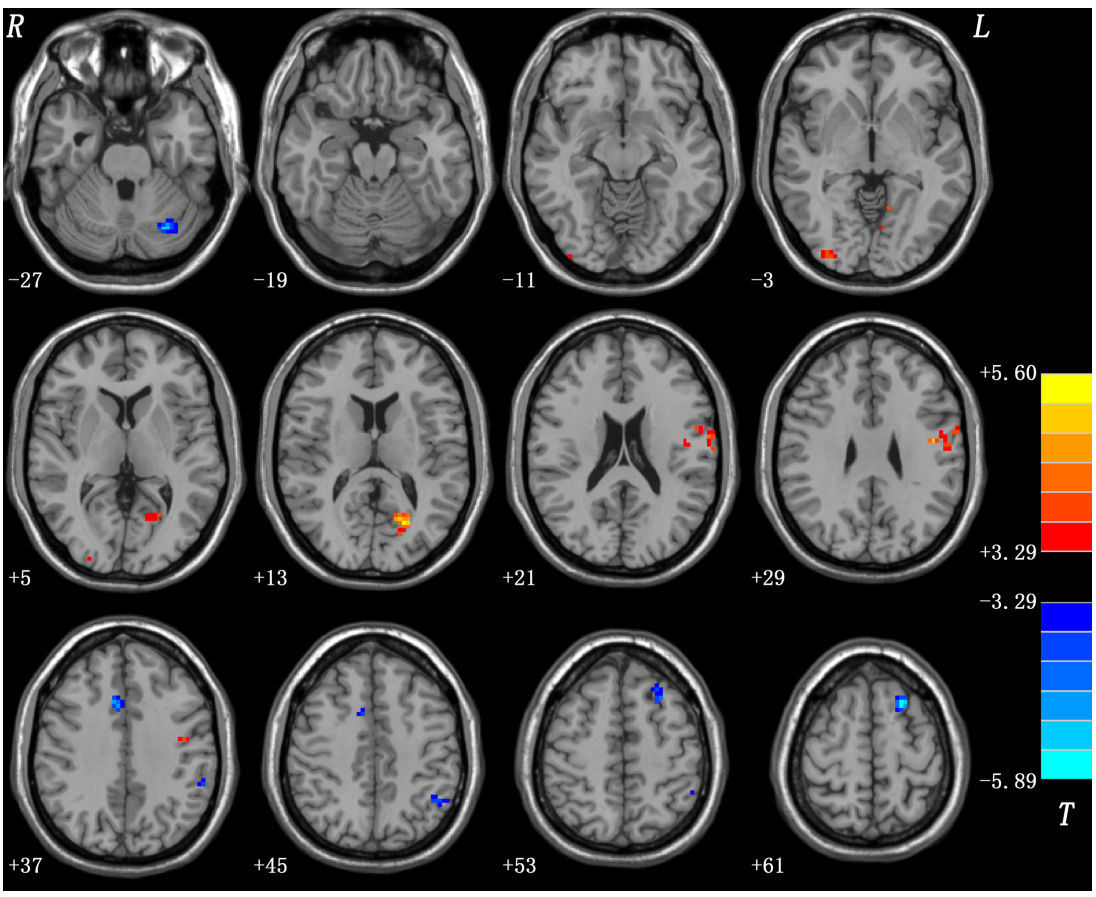


FIGURE 1 | Differences in PAS scores between patients with OCD and HCs. Increased and decreased PAS scores were presented on red and blue colors, respectively. The color bar represents the *t*-values of the group analysis. OCD, obsessive-compulsive disorder; HCs, healthy controls; PAS, parameter of asymmetry.

TABLE 2 | The differences of PAS scores between patients with OCD and HCs.

Cluster location	Peak (MNI)			Number of voxels	*t*-Value
	x	*y*	*z*		
Left posterior cingulate cortex	−21	−66	9	61	5.5996
Left precentral gyrus/postcentral gyrus	−45	−9	30	92	5.1088
Right inferior occipital gyrus	33	−90	−6	30	4.9227
Left dorsolateral prefrontal cortex	−21	15	60	58	−5.8926
Bilateral middle cingulate cortex	9	18	36	42	−4.8529
Left inferior parietal lobule	−48	−54	45	56	−4.0077
Left cerebellum Crus I	−30	−72	−27	38	−5.1077

The significance level was set at $p < 0.05$ for multiple comparisons corrected by Gaussian random field (GRF) theory (voxel significance: $p < 0.001$, cluster significance: $p < 0.05$). Age and the mean FD values were used as covariates to minimize the potential effects of these variables. PAS, parameter of asymmetry; OCD, obsessive-compulsive disorder; HCs, healthy controls.

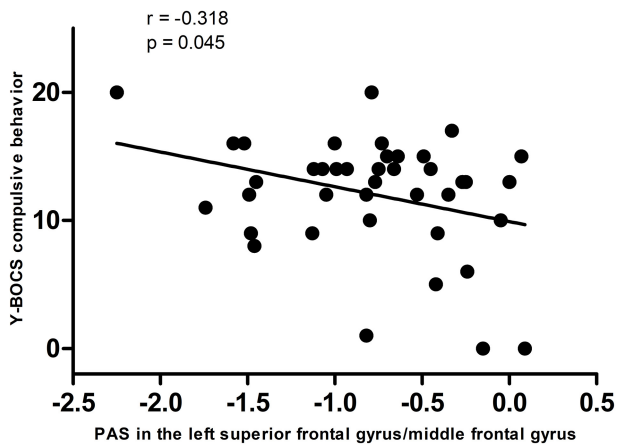


FIGURE 2 | A negative correlation between PAS scores in the left dorsolateral prefrontal cortex and Y-BOCS compulsive behavior scores in the patients. PAS, parameter of asymmetry; Y-BOCS, Yale–Brown Obsessive-Compulsive Scale.

(59/78); 2 = 88.46% (69/78); 3 = 73.08% (57/78); 4 = 74.36% (58/78); 5 = 70.51% (55/78); 6 = 71.79% (56/78); 7 = 79.49% (62/78) (**Figure 3**). The results of SVM analysis showed that the PAS score of the bilateral MCC could be used to discriminate patients with OCD from the HCs with an accuracy of 88.46% (69/78), a sensitivity 87.50% (35/40), and a specificity 89.47% (34/38), respectively (**Figure 4**). A previous study indicated that these indicators above 80% were excellent (Gong et al., 2014). According to this criterion, the SVM result was satisfactory because all indicators were greater than 80%.

DISCUSSION

To the knowledge of the authors, this study was the first to explore the disrupted asymmetries of the inter- and/or intra-hemispheric

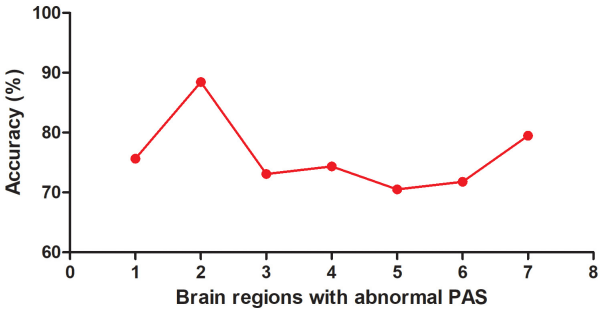


FIGURE 3 | Accuracy (%) of abnormal PAS scores in a single brain region to discriminate OCD from HCs. PAS, parameter of asymmetry; OCD, obsessive-compulsive disorder; HCs, healthy controls; 1, left cerebellum Crus I; 2, bilateral middle cingulate cortex; 3, left dorsolateral prefrontal cortex; 4, right inferior occipital gyrus; 5, left inferior parietal lobule; 6, left posterior cingulate cortex; 7, left precentral gyrus/postcentral gyrus.

FC at rest in medication-free patients with OCD by using the PAS method. Consistent with the hypothesis, this study discovered that patients with OCD showed disrupted asymmetry of the intra-hemispheric FC within (i.e., left DLPFC) and outside (i.e., left PCC, left precentral gyrus/postcentral gyrus, right inferior occipital gyrus, bilateral MCC, left IPL, and left cerebellum Crus I) the CSTC circuits. Decreased PAS in the left DLPFC was negatively correlated with Y-BOCS compulsive behavior scores in OCD. Furthermore, the exploratory SVM results suggested that decreased PAS in the bilateral MCC could discriminate patients with OCD from the HCs with appropriate accuracy, specificity, and sensitivity.

The authors' research and other studies previously found decreased FC within the CSTC circuits at rest in OCD (Wang et al., 2018; Deng et al., 2019; Jia et al., 2020). Although these researchers found that FC strength seemed to exhibit functional asymmetry, they did not explore functional asymmetry of the interhemispheric and/or intra-hemispheric FC at rest in OCD. In the current study, the whole-brain FC was divided into inter- and intra-hemispheric FC, and the results manifested that altered FC in the brain regions within the CSTC circuits (i.e., DLPFC) might be correlated with decreased intra-hemispheric FC within the left hemisphere.

Interestingly, the results showed that decreased PAS scores in the left DLPFC were negatively correlated with the Y-BOCS compulsive behavior scores in patients with OCD. DLPFC has a core role in cognitive flexibility, set shifting, and executive planning (Kwon et al., 2009), which are all damaged in OCD (Menzies et al., 2008). The decreased intra-hemispheric FC in the left DLPFC may be related to the reduced ability to plan activities and the inability to adapt to changes in the external environment in patients with OCD (Chen et al., 2016).

Consistent with the hypothesis, the present study found altered asymmetry of intra-hemispheric FC outside the CSTC circuits in patients with OCD. These brain regions are involved in the default mode network (DMN; i.e., left PCC), executive control network (ECN; i.e., left IPL), sensorimotor network (i.e., left precentral gyrus/postcentral gyrus), and cerebellum

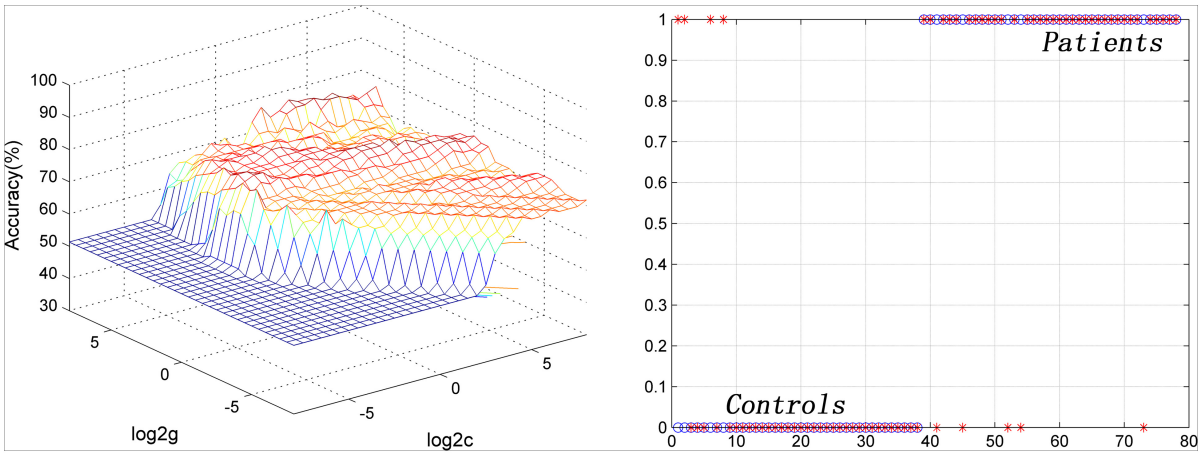


FIGURE 4 | Visualization of SVM classification by using PAS scores in the bilateral middle cingulate cortex to differentiate patients with OCD from HCs. Left: 3D visualization of SVM with the best parameters; right: classification map of the PAS values of the bilateral middle cingulate cortex. SVM, support vector machine; PAS, parameter of asymmetry; OCD, obsessive-compulsive disorder; HCs, healthy controls.

network (i.e., left cerebellum Crus I), which might participate in the pathophysiology of OCD (Cui et al., 2020). Increased FC is regarded as a dedifferentiation and/or compensatory reallocation to the dysfunction (Guo et al., 2015). Therefore, the increased intra-hemispheric FC values in the left PCC and left precentral gyrus/postcentral gyrus in the current research may manifest an effort of compensatory reallocation to the dysfunction in the DMN and sensorimotor network at rest in patients with OCD. Decreased FC is considered as impaired coordination between brain regions (Xu et al., 2019). Thus, the reduced intra-hemispheric FC values in the bilateral MCC, left IPL, and left cerebellum Crus I suggested that the coordination was damaged within the ECN and cerebellum network. The altered intra-hemispheric FC of the DMN, ECN, sensorimotor, and cerebellum network may work together in OCD. These findings provided an additional new evidence for the pathogenesis of OCD.

In the SVM analysis of the current study, feature selection is an important step to reduce the redundancies and select meaningful features from the original feature sets (Zhang et al., 2021). Then, the remaining meaningful features were integrated into a specific classifier in an embedded manner, which was used for SVM training (Zhang et al., 2021). Classification is the approach of classifying the given input by training with an appropriate classifier (Mathew and Anto, 2017). Many researchers suggested that the SVM classifier is one of the best classifiers (Zhang et al., 2011; Menze et al., 2014). In the current study, the results of SVM analysis suggested that reduced intra-hemispheric FC of the bilateral MCC could serve as a potential neurobiological biomarker to discriminate patients with OCD from HCs, with acceptable accuracy, specificity, and sensitivity. MCC plays an essential role in response selection and cognitive task (Shima et al., 1991). Reduced intra-hemispheric FC in the bilateral MCC may be associated with difficulty in response selection between automatic habitual system and cognitive-controlled goal-directed system in patients with OCD (Dong et al., 2020).

Several innovative aspects of the current study must be mentioned. First, the division of the whole-brain FC into interhemispheric FC and intra-hemispheric FC may help explain whether abnormal FC resulted from interhemispheric FC and/or intra-hemispheric FC at rest in OCD. Second, an exploratory SVM analysis was used to investigate whether FC asymmetry could be used as a biological marker to discriminate OCD from HCs at the individual level. Third, medication-free patients with OCD were recruited, which may eliminate the potential effects of the heterogeneity of diseases and drugs on the brain FC at rest.

However, several limitations still exist. First, white and gray matters were not assessed, and their potential effects on interhemispheric FC and/or intra-hemispheric FC were unclear. Second, patients with OCD were not classified into subtypes in accordance with their clinical symptoms.

CONCLUSION

In summary, the study showed disrupted asymmetry of the intra-hemispheric FC within and outside the CSTC circuits at rest in medication-free patients with OCD. The reduced intra-hemispheric FC in the bilateral MCC could be used as a biomarker to distinguish individuals with OCD from HCs. Future studies should address these limitations and further explore the clinical utility of using altered asymmetry of the intra-hemispheric FC for differentiation in OCD and prediction of treatment response (i.e., medicine, psychotherapy, and neuroregulation) in patients with OCD.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by this study was approved by the Research Ethics Committee of Qiqihar Medical University. The participants were informed of the study procedures, and they signed a written informed consent. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CJ, YO, and YC engaged in data analysis and wrote the manuscript. PL and WG designed and conducted the study. JM, CZ, DL, RY, TS, LS, YW, GZ, ZS, WW, and XW participated in patient assessment and imaging data collection. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Project of Heilongjiang Provincial Department of Education (2018-KYYWF-0111).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.634557/full#supplementary-material

REFERENCES

Boedhoe, P. S., Schmaal, L., Abe, Y., Ameis, S. H., Arnold, P. D., Batistuzzo, M. C., et al. (2017). Distinct subcortical volume alterations in pediatric and adult OCD: a worldwide meta- and mega-analysis. *Am. J. Psychiatry* 174, 60–69. doi: 10.1176/appi.ajp.2016.16020201

Bu, X., Hu, X., Zhang, L., Li, B., Zhou, M., Lu, L., et al. (2019). Investigating the predictive value of different resting-state functional MRI parameters in obsessive-compulsive disorder. *Transl. Psychiatry* 9:17. doi: 10.1038/s41398-018-0362-9

Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., et al. (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J. Neurosci.* 29, 1860–1873. doi: 10.1523/JNEUROSCI.5062-08.2009

Chen, Y., Meng, X., Hu, Q., Cui, H., Ding, Y., Kang, L., et al. (2016). Altered resting-state functional organization within the central executive network in obsessive-compulsive disorder. *Psychiatry Clin. Neurosci.* 70, 448–456. doi: 10.1111/pcn.12419

Crow, T. J., Ball, J., Bloom, S. R., Brown, R., Bruton, C. J., Colter, N., et al. (1989). Schizophrenia as an anomaly of development of cerebral asymmetry. a postmortem study and a proposal concerning the genetic basis ofthe disease. *Arch. Gen. Psychiatry* 46, 1145–1150. doi: 10.1001/archpsyc.1989.01810120087013

Cui, G., Ou, Y., Chen, Y., Lv, D., Jia, C., Zhong, Z., et al. (2020). Altered global brain functional connectivity in drug-naive patients with obsessive-compulsive disorder. *Front. Psychiatry* 11:98. doi: 10.3389/fpsyt.2020.00098

Deng, K., Qi, T., Xu, J., Jiang, L., Zhang, F., Dai, N., et al. (2019). Reduced interhemispheric functional connectivity in obsessive-compulsive disorder patients. *Front. Psychiatry* 10:418. doi: 10.3389/fpsyt.2019.00418

Ding, Y. D., Yang, R., Yan, C. G., Chen, X., Bai, T. J., Bo, Q. J., et al. (2021). Disrupted hemispheric connectivity specialization in patients with major depressive disorder: evidence from the REST-meta-MDD Project. *J. Affect. Disord.* 284, 217–228. doi: 10.1016/j.jad.2021.02.030

Dong, C., Yang, Q., Liang, J., Seger, C. A., Han, H., Ning, Y., et al. (2020). Impairment in the goal-directed corticostriatal learning system as a biomarker for obsessive-compulsive disorder. *Psychol. Med.* 50, 1490–1500. doi: 10.1017/S0033291719001429

Doron, K. W., and Gazzaniga, M. S. (2008). Neuroimaging techniques offer new perspectives on callosal transfer and interhemispheric communication. *Cortex* 44, 1023–1029. doi: 10.1016/j.cortex.2008.03.007

Federspiel, A., Begre, S., Kiefer, C., Schroth, G., Strik, W. K., and Dierks, T. (2006). Alterations of white matter connectivity in first episode schizophrenia. *Neurobiol. Dis.* 22, 702–709. doi: 10.1016/j.nbd.2006.01.015

First, M. B., Spitzer, R. L., Gibbon, M., and Williams, J. B. W. (1996). *Structured Clinical Interview for DSMIV Axis I Disorders. Patient Edition*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute.

First, M. B., Spizer, R. L., Gibbon, M., and Williams, J. B. W. (2001). *Structures Clinical Interview for DSM-IV-TR Axis I Disorder Version. Non-patient edition (SCID-I/NP)*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute.

Gong, Q., Li, L., Tognin, S., Wu, Q., Pettersson-Yeo, W., Lui, S., et al. (2014). Using structural neuroanatomy to identify trauma survivors with and without post-traumatic stress disorder at the individual level. *Psychol. Med.* 44, 195–203. doi: 10.1017/s0033291713000561

Guo, W., Liu, F., Xiao, C., Liu, J., Yu, M., Zhang, Z., et al. (2015). Increased short-range and long-range functional connectivity in first-episode, medication-naive schizophrenia at rest. *Schizophr. Res.* 166, 144–150. doi: 10.1016/j.schres.2015.04.034

Herbert, M. R., Ziegler, D. A., Deutsch, C. K., O'Brien, L. M., Kennedy, D. N., Filipek, P. A., et al. (2005). Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. *Brain* 128(Pt 1), 213–226. doi: 10.1093/brain/awh330

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., and Xu, W. (2018). Applications of Support Vector Machine (SVM) learning in cancer genomics. *Cancer Genom. Proteom.* 15, 41–51. doi: 10.21873/cgp.20063

Jia, C., Ou, Y., Chen, Y., Li, P., Lv, D., Yang, R., et al. (2020). Decreased resting-state interhemispheric functional connectivity in medication-free obsessive-compulsive disorder. *Front. Psychiatry* 11:559729. doi: 10.3389/fpsyt.2020.559729

Kong, X. Z., Boedhoe, P. S. W., Abe, Y., Alonso, P., Ameis, S. H., Arnold, P. D., et al. (2020). Mapping cortical and subcortical asymmetry in obsessive-compulsive disorder: findings from the ENIGMA consortium. *Biol. Psychiatry* 87, 1022–1034. doi: 10.1016/j.biopsych.2019.04.022

Kwon, J. S., Jang, J. H., Choi, J. S., and Kang, D. H. (2009). Neuroimaging in obsessive-compulsive disorder. *Expert Rev. Neurother.* 9, 255–269. doi: 10.1586/14737175.9.2.255

Liu, D., Qian, H., Dai, G., and Zhang, Z. (2013). An iterative SVM approach to feature selection and classification in high-dimensional datasets. *Pattern Recogn.* 46, 2531–2537. doi: 10.1016/j.patcog.2013.02.007

Liu, F., Zhu, C., Wang, Y., Guo, W., Li, M., Wang, W., et al. (2015). Disrupted cortical hubs in functional brain networks in social anxiety disorder. *Clin. Neurophysiol.* 126, 1711–1716. doi: 10.1016/j.clinph.2014.11.014

Mathew, A. R., and Anto, P. B. (2017). ""Tumor detection and classification of MRI brain image using wavelet transform and SVM"," in *Proceedings of the 2017 International Conference on Signal Processing and Communication (ICSPC)*, (Piscataway, NJ: IEEE), 75–78. doi: 10.1109/CSPC.2017.8305810

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* 34, 1993–2024. doi: 10.1109/TMI.2014.2377694

Menzies, L., Chamberlain, S. R., Laird, A. R., Thelen, S. M., Sahakian, B. J., and Bullmore, E. T. (2008). Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: the orbitofronto-striatal model revisited. *Neurosci. Biobehav. Rev.* 32, 525–549. doi: 10.1016/j.neubiorev.2007.09.005

Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., and Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological

and psychiatric disease: a critical review. *Neurosci. Biobehav. R.* 36, 1140–1152. doi: 10.1016/j.neubiorev.2012.01.004

Peng, Z., Li, G., Shi, F., Shi, C., Yang, Q., Chan, R. C., et al. (2015). Cortical asymmetries in unaffected siblings of patients with obsessive-compulsive disorder. *Psychiatry Res.* 234, 346–351. doi: 10.1016/j.pscychresns.2015.10.005

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Samara, A., and Tsangaris, G. T. (2011). Brain asymmetry: both sides of the story. *Expert Rev. Proteomics* 8, 693–703. doi: 10.1586/epr.11.62

Schüz, A., and Preissl, H. (1996). Basic connectivity of the cerebral cortex and some considerations on the corpus callosum. *Neurosci. Biobehav. Rev.* 20, 567–570. doi: 10.1016/0149-7634(95)00069-0

Shaw, P., Lalonde, F., Lepage, C., Rabin, C., Eckstrand, K., Sharp, W., et al. (2009). Development of cortical asymmetry in typically developing children and its disruption in attention-deficit/hyperactivity disorder. *Arch. Gen. Psychiat.* 66, 888–896. doi: 10.1001/archgenpsychiatry.2009.103

Shima, K., Aya, K., Mushiake, H., Inase, M., Aizawa, H., and Tanji, J. (1991). Two movementrelated foci in the primate cingulate cortex observed in signal-triggered and self-paced forelimb movements. *J. Neurophysiol.* 65, 188–202. doi: 10.1152/jn.1991.65.2.188

Su, Q., Yu, M., Liu, F., Li, Y., Li, D., Deng, M., et al. (2020). Abnormal functional asymmetry in the salience and auditory networks in first-episode, drug-naive somatization disorder. *Neuroscience* 444, 1–8. doi: 10.1016/j.neuroscience.2020.07.043

Wang, L., Dai, Z., Peng, H., Tan, L., Ding, Y., He, Z., et al. (2014). Overlapping and segregated resting-state functional connectivity in patients with major depressive disorder with and without childhood neglect. *Hum. Brain Mapp.* 35, 1154–1166. doi: 10.1002/hbm.22241

Wang, Y., Wang, Y. M., Lui, S. S. Y., and Chan, R. C. K. (2020). Clinical implication of brain asymmetries in psychiatric disorders. *Biol. Psychiatry* 87, 1014–1016. doi: 10.1016/j.biopsych.2020.04.006

Wang, Y. H., Chen, Y. H., Li, S. F., Lv, D., Zhao, A. M., Meng, X., et al. (2018). Reduced interhemispheric resting-state functional homotopy in obsessive-compulsive disorder. *Neuropsychiatry-Lond* 8, 1038–1045.

Xu, T., Zhao, Q., Wang, P., Fan, Q., Chen, J., Zhang, H., et al. (2019). Altered resting-state cerebellar-cerebral functional connectivity in obsessive-compulsive disorder. *Psychol. Med.* 49, 1156–1165. doi: 10.1017/S0033291718001915

Yan, C. G., Wang, X. D., Zuo, X. N., and Zang, Y. F. (2016). DPABI: data processing & analysis for (Resting-State) brain imaging. *Neuroinformatics* 14, 339–351. doi: 10.1007/s12021-016-9299-4

Zach, P., Valeš, K., Stuchlik, A., Čermáková, P., Mrzílková, J., Koutella, A., et al. (2016). Effect of stress on structural brain asymmetry. *Neuro. Endocrinol. Lett.* 37, 253–264.

Zhang, N., Ruan, S., Lebonvallet, S., Liao, Q., and Zhu, Y. (2011). Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation. *Comput. Vis. Image Und.* 115, 256–269. doi: 10.1016/j.cviu.2010.09.007

Zhang, Y., Wang, S., Xia, K., Jiang, Y., and Qian, P. (2021). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inform. Fusion* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002

Zhu, F., Liu, F., Guo, W., Chen, J., Su, Q., Zhang, Z., et al. (2018). Disrupted asymmetry of inter- and intra-hemispheric functional connectivity in patients with drug-naive, first-episode schizophrenia and their unaffected siblings. *EBioMedicine* 36, 429–435. doi: 10.1016/j.ebiom.2018.09.012

Zhu, F., Liu, Y., Liu, F., Yang, R., Li, H., Chen, J., et al. (2019). Functional asymmetry of thalamocortical networks in subjects at ultra-high risk for psychosis and first-episode schizophrenia. *Eur. Neuropsychopharmacol.* 29, 519–528. doi: 10.1016/j.euroneuro.2019.02.006

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 17 June 2021
doi: 10.3389/fnins.2021.670287

DeepNavNet: Automated Landmark Localization for Neuronavigation

*Christine A. Edwards[1,2,3], Abhinav Goyal[2,4], Aaron E. Rusheen[2,4], Abbas Z. Kouzani[1] and Kendall H. Lee[2,3,4,5]**

[1] School of Engineering, Deakin University, Geelong, VIC, Australia, [2] Department of Neurologic Surgery, Mayo Clinic, Rochester, MN, United States, [3] Mayo Clinic Graduate School of Biomedical Sciences, Mayo Clinic, Rochester, MN, United States, [4] Mayo Clinic College of Medical Scientist Training Program, Mayo Clinic, Rochester, MN, United States, [5] Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN, United States

Functional neurosurgery requires neuroimaging technologies that enable precise navigation to targeted structures. Insufficient image resolution of deep brain structures necessitates alignment to a brain atlas to indirectly locate targets within preoperative magnetic resonance imaging (MRI) scans. Indirect targeting through atlas-image registration is innately imprecise, increases preoperative planning time, and requires manual identification of anterior and posterior commissure (AC and PC) reference landmarks which is subject to human error. As such, we created a deep learning-based pipeline that consistently and automatically locates, with submillimeter accuracy, the AC and PC anatomical landmarks within MRI volumes without the need for an atlas. Our novel deep learning pipeline (DeepNavNet) regresses from MRI scans to heatmap volumes centered on AC and PC anatomical landmarks to extract their three-dimensional coordinates with submillimeter accuracy. We collated and manually labeled the location of AC and PC points in 1128 publicly available MRI volumes used for training, validation, and inference experiments. Instantiations of our DeepNavNet architecture, as well as a baseline model for reference, were evaluated based on the average 3D localization errors for the AC and PC points across 311 MRI volumes. Our DeepNavNet model significantly outperformed a baseline and achieved a mean 3D localization error of 0.79 ± 0.33 mm and 0.78 ± 0.33 mm between the ground truth and the detected AC and PC points, respectively. In conclusion, the DeepNavNet model pipeline provides submillimeter accuracy for localizing AC and PC anatomical landmarks in MRI volumes, enabling improved surgical efficiency and accuracy.

Keywords: deep brain stimulation, deep learning, human-machine teaming, landmark localization, neuroimaging, neuronavigation, neurosurgery planning

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Jon Orlando Cleary, Guy's and St Thomas' NHS Foundation Trust, United Kingdom
David Haynor, University of Washington, United States

***Correspondence:**
Kendall H. Lee
lee.kendall@mayo.edu

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *20 February 2021*
Accepted: *25 May 2021*
Published: *17 June 2021*

Citation:
Edwards CA, Goyal A, Rusheen AE, Kouzani AZ and Lee KH (2021) DeepNavNet: Automated Landmark Localization for Neuronavigation. Front. Neurosci. 15:670287. doi: 10.3389/fnins.2021.670287

INTRODUCTION

Imaging technologies such as magnetic resonance imaging (MRI) provide visualization of brain structures that enable neurosurgeons to plan accurate and safe surgical trajectories (Edwards et al., 2017). In MRI-guided functional neurosurgery, image space is registered to a stereotactic head frame with a built-in three-dimensional (3D) coordinate system using stereotactic planning

software (Edwards et al., 2018). This software allows surgeons to view the neuroimaging data to derive 3D coordinates of the brain target(s) to plan the surgical trajectory path(s). This approach has been predominantly used to implant electrodes for deep brain stimulation and to deliver focused ultrasonic ablation, both of which provide therapeutic relief of debilitating movement and psychiatric disorders (Elias et al., 2016; Edwards et al., 2017, 2018).

Although MRI technologies (1.5 and 3 Tesla) have improved over the last decade, visibility of deep intracranial target nuclei remains poor due to imaging protocols that do not provide requisite image resolution and contrast. As such, it is standard practice to use "indirect" targeting methods to locate DBS targets in preoperative MRI scans. DBS targets are typically indirectly identified by manually locating visible anatomical landmarks, i.e., anterior and posterior commissure points (AC and PC), and using these reference points to align the patient's preoperative MRI scan to a stereotactic brain atlas. In contrast, "direct" targeting methods may be used when targeted structures are visible and directly identifiable in preoperative images, e.g., acquired using special MRI sequences designed to target specific structures, acquired in an ultra-high field MRI scanner, or in visualizations created through multimodal fusion techniques (Cho et al., 2010; Grewal et al., 2018, 2020; Hartmann et al., 2019). Despite advances in visualizations and direct targeting methods, manual localization of AC and PC points remain standard practice and ACPC alignment is valuable beyond targeting such as for multimodal visualizations for pre- and post-operative assessments (Isaacs et al., 2020). The locations of deep brain surgical targets are estimated by defining a distance from the mid AC-PC point and crosschecked against the registered atlas. Targets include various neurosurgical structures including the subthalamic nucleus (STN) for Parkinson's disease, ventral intermediate nucleus for essential tremor, globus pallidus internus for dystonia, and the centromedian-parafascicular complex for Tourette's syndrome. For example, the STN is typically located 5 mm inferior, 1–2 mm posterior, and 9–12 mm lateral from the mid AC-PC point (Slavin et al., 2006). Therefore, accurate identification of the AC and PC points is of vital importance for safe and efficacious outcomes.

A review of factors impacting DBS targeting accuracy indicate that human errors during image-guided planning contribute to an overall average localization error of 1–2 mm (Li et al., 2016). Error in the localization of the AC and PC anatomical landmarks has contributed to misidentification of the target nuclei and misalignment with the stereotactic brain atlas, leading to surgical error (Pallavaram et al., 2008). As a secondary verification, intraoperative electrophysiologic microelectrode recording is performed. This method increases the risk of adverse events such as intracerebral hemorrhage and seizures, especially if initial AC-PC determinations are incorrect (Hariz and Fodstad, 1999; Binder et al., 2005). Thus, a novel system that provides more accurate and reliable AC-PC localization would improve functional neurosurgery. With the revival of artificial intelligence (AI), sparked by deep learning methods demonstrating human-level performance in computer vision applications, it is now possible to conceive an AI-powered DBS surgical targeting system that augments surgical planning by the neurosurgery team (Panesar et al., 2020).

Early methods to automatically localize anatomical landmarks relied on the identification of surrounding anatomical structures, e.g., ventricles, corpus callosum and fornix. These methods used low-level image processing techniques (e.g., edge detection, histogram analysis, and region growing) coupled with anatomical knowledge-based rules to conduct scene analysis (Verard et al., 1997; Bhanu Prakash et al., 2006). Furthermore, they depend on intensity-based image processing techniques (e.g., binarization) to accurately identify regions of interest, and are thus considered antiquated in light of more modern robust methods. These traditional image processing approaches were superseded by atlas registration and machine learning based techniques (D'Haese et al., 2005; Pallavaram et al., 2009). Ardekani and Bachmann introduced a template model to characterize the displacement of the AC and PC points from an additional more prominent anatomical landmark (i.e., midbrain-pons junction), and demonstrated increased accuracy over methods that relied on primitive image processing techniques (Ardekani and Bachman, 2009). A random forest regression approach demonstrated increased accuracy from millimeters to submillimeter localization errors compared to atlas-based methods for detecting AC and PC points within a limited set of 100 T1-weighted MR volumes acquired uniformly from an institution (Liu and Dawant, 2015). These regression forests were built upon texture-based features (i.e., local binary patterns), which have largely been replaced by more robust deeply-learned features in the computer vision community (LeCun et al., 2015).

Since the resurgence of artificial intelligence in the past decade, deep learning based methods continue to rapidly evolve and dominate computer vision applications for natural images and videos. Deep regression approaches are commonly used for facial landmark detection, human pose estimation and image registration applications (Lathuiliere et al., 2019). Although deep learning methods applied in the medical imaging domain were lagging behind, they are gaining momentum (Akkus et al., 2017; Hinton, 2018; Korfiatis and Erickson, 2019). Variations of fully convolutional network (FCN) deep learning architectures (e.g., U-Net) have achieved state-of-the-art results for image segmentation and localization within medical imaging (Falk et al., 2019). Payer et al. augmented an FCN architecture with a spatial configuration network (SCN) module which preserves the geometric relationships between landmarks, and enables more accurate heatmap regression to localize landmarks where large annotated medical image datasets are less readily available (Payer et al., 2016, 2019). Zhang et al. (2017) proposed a cascaded two-stage multi-task deep convolutional neural network architecture to simultaneously localize hundreds of anatomical landmarks in medical images where a limited amount of training data is available (Zhang et al., 2017). In particular, Zhang et al. detected 1200 landmarks in 700 MRI scans, with a mean error of 2.96 mm (Zhang et al., 2017). Given the achievements of using deep learning methods for challenging neuroimaging tasks, it is expected that deep learning approaches would achieve increased accuracy for the presumably less complex task of localizing midline points, e.g., the AC and PC points.

The aforementioned deep learning based methods have demonstrated state-of-the-art and promising results for a variety of medical image modalities and applications. Even so, these methods neither report submillimeter accuracy in localization of landmarks in neuroimaging data, nor report the localization of AC and PC points as required for targeted functional neurosurgery planning. Here, we present the development and validation of a novel deep learning-based heatmap regression pipeline that is built upon an implementation of a 3D residual network (Liu et al., 2018) that demonstrates submillimeter accuracy in localization of the AC and PC landmarks within MRI volumes.

MATERIALS AND METHODS

Data Preparation

Experiments were conducted using two datasets collated from publicly available T1-weighted 1.5-Tesla (T) MRI volumes, which we will refer to as ACPC-MRI-1 and ACPC-MRI-2. Each MRI scan is 256 × 256 × 150 voxels with a cubic millimeter resolution per voxel. The first dataset, ACPC-MRI-1, is comprised of 908 MRI volumes across 687 subjects from the following publicly available data: Open Access Series of Imaging Studies (OASIS) (Marcus et al., 2007) and Mindboggle-101 (Klein et al., 2017). The OASIS dataset is comprised of scans for subjects aged 18 to 96 years, with a subset of subjects aged 60 years and older who were diagnosed with the onset of dementia to moderate Alzheimer's disease (Marcus et al., 2007). Two to four additional scans are included for the subset of 100 subjects with dementia and 20 healthy subjects (Marcus et al., 2007). The MindBoggle dataset includes T1-weighted MRI volumes from 101 healthy subjects (Klein and Tourville, 2012). The second dataset, ACPC-MRI-2, contains 220 volumes across 158 subjects from the OASIS-3 dataset (LaMontagne et al., 2019). The AC and PC anatomical landmarks were manually localized for a total of 1,128 MRI annotated volumes across ACPC-MRI-1 and ACPC-MRI-2 datasets.

An annotation protocol was established to label the AC and PC anatomical landmarks for each T1-weighted MRI volume. To streamline the annotation process, each volume was aligned to a common MRI template using a rigid registration approach, which is publicly available in the Advanced Normalization Tools (ANTs) software toolkit (Avants et al., 2011). Each normalized MRI volume was imported into the publicly available 3D Slicer Tool[1] and viewed with a predefined AC and PC template overlaid on the volume (Kikinis et al., 2014; Fedorov et al., 2016). Using the 3D Slicer Tool, each annotator manually fine-tuned reference point locations to the posterior and anterior edges of the AC and PC points, respectively, and saved the 3D coordinates for each point to a fiducial markup file per MRI volume (see **Figure 1**). Without fine tuning the locations, the average 3D Euclidean distances between initial locations provided by our template and our ground truth locations of AC and PC points were 6.91 ± 5.50 mm and 6.44 ± 4.89 mm (N = 1128), respectively. For the ACPC-Dataset-1, a team of five annotators ranging from novice to an experienced neurosurgeon labeled the AC and PC points for a total of 908 T1-weighted MRI volumes across 687 subjects. Multiple annotators labeled a subset of the MRI scans (N = 274), and the resulting 3D coordinates for each landmark were averaged and used as ground truth. For quality control, when multiple labels were available for a given point, the 3D Euclidean distance between points was calculated The average AC and PC distance, across annotators, was 1.14 ± 0.57 mm (2.92 mm max error) and 1.04 ± 0.53 mm (2.94 mm max error), respectively. For distances greater than 2.0 mm, the labels were visually evaluated and adjusted by an expert annotator. For MRI volumes where only a single annotation was available and the annotator was considered a novice, the label was visually inspected and adjusted as needed. Approximately 10% of the labeled points were re-adjusted manually within the Slicer Tool by an expert-level annotator. For ACPC-Dataset-2, an expert-level annotator labeled the AC and PC landmarks of all the volumes, using the same annotation protocol described for ACPC-Dataset-1.

The DeepNavNet models described in this paper regress from an input MRI volume to an output heatmap volume with spheres centered around 3D coordinates of the AC and PC landmarks within each volume. Therefore, training data preparation included an additional step of creating heatmap volumes with Gaussian spheres, with sigma set to 3, centered around 3D coordinates of the AC and PC landmarks within each MRI volume. Each heatmap was created with the same dimensions as their corresponding MRI volume, and the voxel intensity values were normalized to a maximum value of 1. The ACPC-MRI-1 dataset was used for training, validation and testing, whereas the ACPC-MRI-2 data was used for additional testing only. As such, heatmap creation was only necessary for the ACPC-MRI-1 dataset.

Deep Learning Software Framework and Computational Resources

All experiments were conducted using the open source NiftyNet (0.6.0) framework for medical image analysis built on Google's TensorFlow (1.15.0) deep learning framework, on a Lambda Labs TensorBook laptop with an NVIDIA GeForce GTX 1070 with 32 GB of memory (Gibson et al., 2018). DeepNavNet's localization model was built upon NiftyNet's regression application, the original implementation of which utilize an iterative sampling method to regress from MRI T2-weighted images to CT images (Berger et al., 2017). To create DeepNavNet, this application was adapted to train models to learn a mapping from MRI volumes to volumetric heatmaps concentrated around corresponding AC and PC anatomical landmark locations. NiftyNet configuration files were modified to experiment with different training paradigms to create an optimized model to automatically locate the AC and PC with the greatest accuracy, and a post-processing Python script was created to extract the 3D coordinates for the AC and PC landmarks from heatmap volumes.

[1] 3D Slicer Tool available from http://www.slicer.org (accessed May 2018).

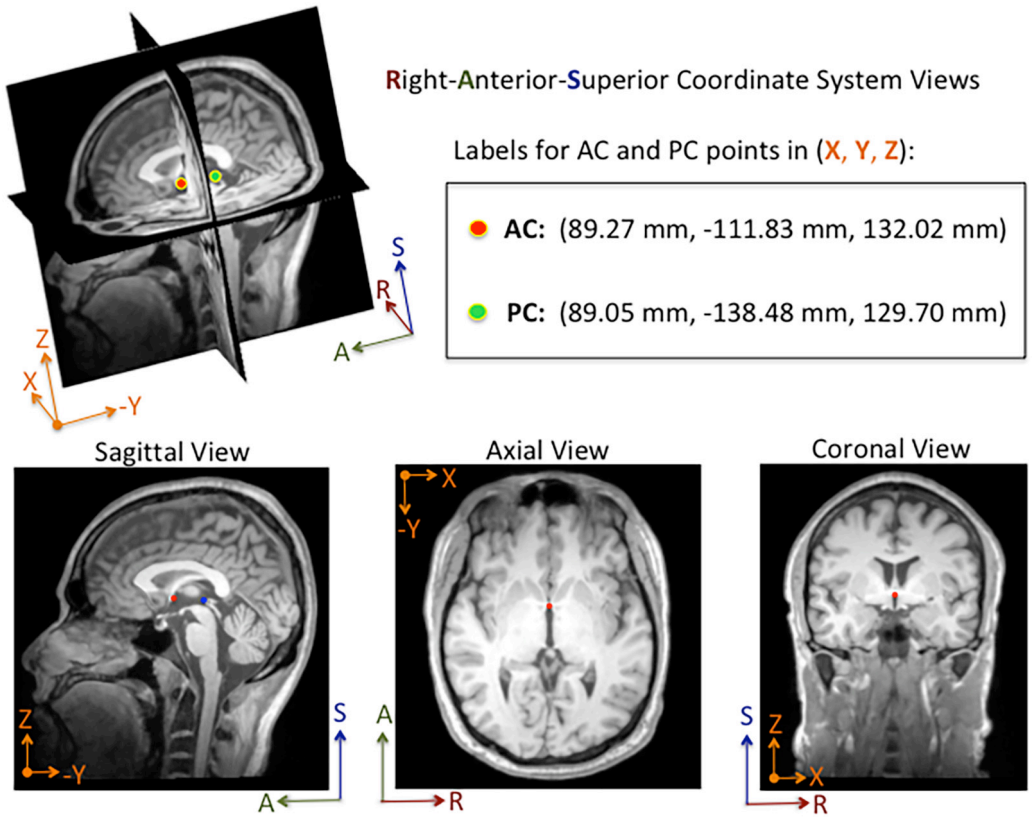


FIGURE 1 | Example of an MRI volume annotated using 3D Slicer software. The volume is displayed in Right-Anterior-Superior (RAS) coordinate space. The annotator manually adjusts fiducial markers overlaid on the AC and PC points from linked 2-D views. For this subject, the AC and PC points are both visible in the same sagittal slice, whereas only the AC point is visible in the axial and coronal slices shown here. Depending on the orientation of the brain within the MRI volume, the AC and PC points are sometimes both visible within a single sagittal and an axial slice. The 3D coordinates for each point are displayed in millimeters in the Cartesian X-Y-Z space. The labels and corresponding 3D coordinates are saved to a fiducial comma-separated values (.fcsv) file.

Deep Learning Architecture and Training Experiments

DeepNavNet models were trained using a 3D residual network architecture that was originally created to segment volumetric medical images, with demonstrated success parcellating anatomical structures within neuroimaging data (Li et al., 2017). This 20-layered network includes dilation to widen the receptive field to provide voxel context and more spatial resolution compared to other convolutional neural networks used for segmentation applications, e.g., (U-Net) (Ronneberger et al., 2015; Li et al., 2017). Also, the network includes residual connections, which are known to enable effective training of deep networks by preserving and propagating information through the networks (Lundervold and Lundervold, 2019). All experiments used NiftyNet's built-in implementation of this architecture, which is referred to as *highres3dnet_large* in the software framework.

NiftyNet's framework includes a patch-based analysis pipeline with different sampling methods to extract windows from input data, as is often required for neuroimaging deep learning applications to counter challenges of high-dimensional data and GPU memory limitations. DeepNavNet builds upon this feature to employ a weighted sampling method with Gaussian sphere heatmaps as sampling priors for the AC and PC locations. This heatmap was created from the initial AC and PC locations in the template used during the manual annotation process previously described. Also, this template heatmap was used as an error map for regression. NiftyNet's implementation includes an option to fine tune sampling by updating the error maps, however, this was not necessary for our application. DeepNavNet was trained with a spatial window size set to (72, 72, 1) and a random flipping layer set to augment the training data with three orthogonal 2D views for patch analysis. In addition to flipping sampled windows, training data was randomly augmented by up to 10% increases and decreases in spatial volume scaling. For all experiments, the training batch size was set to 64, which is the number of windows processed per iteration.

To optimize DeepNavNet for automated AC and PC localization, we generated variations using a combination of different training parameters. Loss functions used include the root mean squared error (RMSE) and Huber loss functions (Lathuiliere et al., 2019). Regularization was achieved by using a combination of data augmentation with lasso regression (L1) or ridge regression (L2) regularization methods. Adaptive learning

rate techniques used were adaptive moment estimation (Adam) and root mean square propagation (RMSprop), with initial learning rate and weight decay parameters set to 0.005 and 1e-5, respectively (Kingma and Ba, 2015). Each variation of the DeepNavNet model was trained for 2,500 iterations and corresponding learning curves were captured via Tensorboard software. An 80:10:10 ratio was used to separate the ACPC-MRI-1 dataset into subsets of data used for training (N = 726), validation (N = 91) and testing (N = 91). **Table 1** provides a list of training experiments that created 8 variations of the DeepNavNet model, which were then tested for accuracy during inference experiments (see **Figure 2**). **Figure 2** also lists the training hyperparameters employed within DeepNavNet, which were optimized during model validation.

Inference Experiments

Inference experiments were conducted to evaluate model accuracy as measured by calculating the 3D Euclidean distance error between the ground truth and predicted AC and PC coordinates. Initial evaluations of all the models used a test set of 91 volumes from ACPC-MRI-1 dataset, which were volumes not used to train nor validate the models during the training process. If the model was able to detect the AC and PC coordinates with reasonable accuracy, then additional testing was performed with the full-set of 220 volumes from ACPC-MRI-2 dataset, and the resulting 3D Euclidean localization errors were combined to determine the overall accuracy across both datasets (N = 311). First, a baseline evaluation was established using the publicly available Automatic Registration Toolkit (ART) software package, which includes an Ardekani and Bachmann model-based approach to detect AC and PC points and a model trained on T1-weighted MRI volumes (Ardekani and Bachman, 2009) (Klein et al., 2009). The acpcdetect module within ART was used to detect the AC and PC locations within the combined dataset of 311 MRI test volumes. A Python script was created to extract the AC and PC 3D coordinates from the output text file per MRI volume.

Using NiftyNet software, inference experiments were conducted to test the accuracy of each instantiation of the DeepNavNet model listed in **Table 1**. The DeepNavNet model predicts a heatmap for each input MRI volume (see **Figure 3**).

TABLE 1 | DeepNavNet training experiments.

DeepNavNet Model #	Loss function		Regularization		Adaptive learning rate	
	RMSE	Huber	L1	L2	Adam	RMSProp
1	✓		✓			✓
2	✓			✓		✓
3		✓	✓			✓
4		✓		✓		✓
5	✓		✓		✓	
6	✓			✓	✓	
7		✓	✓		✓	
8		✓		✓	✓	

A post-processing Python script was created to extract the predicted 3D coordinates for the AC and PC landmarks from the predicted heatmaps. Since input MRI volumes are aligned to a template volume, the AC and PC landmarks are expected to be relatively centered within a region of the volume. Furthermore, there are anatomical constraints regarding the position of the AC and PC with respect to each other. As such, post-processing included applying a binary mask to filter out false positives outside this region of interest, applying a Gaussian filter to smooth the spheres, and assigning AC and PC labels to maximum voxels identified based upon their spatial relationship to each other (see **Figure 4**).

Statistical Analysis

Outliers were identified and removed from each model's localization errors prior to comparison between models to ensure fair comparison. Outliers were defined as those errors which were greater than 3 scaled median absolute deviations (MAD) away from the median of the errors for that model. Scaled $MAD = c * median(|A_i - median(A)|)$ for all i in the model, where $c = \frac{-1}{\sqrt{2} * erfc^{-1}(1.5)}$. After outliers were removed, AC errors across models were compared with a one-way analysis of variance (ANOVA), and post-hoc comparisons between DeepNavNet models and baseline were conducted with one-tailed unpaired t-tests, while post-hoc comparisons between DeepNavNet models were done with two-tailed unpaired t-tests. A Bonferroni correction with 6! = 720 comparisons was applied to each significance level (α = 0.05) to correct for multiple comparisons. P-values less than 0.001 are reported as such rather than reporting the exact value. The same procedure was performed for PC localization errors across all models.

RESULTS

Our best performing DeepNavNet (Model-1) achieved a mean localization error of 0.79 $\pm$ 0.33 mm (1.62 mm max error) and 0.78 $\pm$ 0.33 (1.66 mm max error) between the ground truth and the detected AC and PC points across N = 311 volumes, respectively (see **Table 2**). A one-way ANOVA was conducted across all models separately for AC and PC localization errors to assess differences in model ability to accurately localize these landmarks. We found that models did perform differently for both AC and PC localization (AC: F = 47, $p << 0.001$; PC: F = 19, $p << 0.001$). Each DeepNavNet instantiation was then compared directly to the baseline model, separately for AC and PC localization error. Models 1-6 were found to significantly outperform the baseline model for accurately localizing the AC landmark (one-tailed unpaired t-tests, all adjusted $p << 0.001$; see **Figure 5A**). For localizing the PC landmark, models 1, 2, and 5 were found to significantly outperform the baseline model (one-tailed unpaired t-tests, all adjusted $p << 0.001$; see **Figure 5B**).

To assess model precision, we assessed the spread of AC and PC localization error distributions across our models with outliers kept in the data. These outliers represent failures of the model to reliably locate the AC and PC landmarks across trials. The number of outliers and the standard deviation of the models

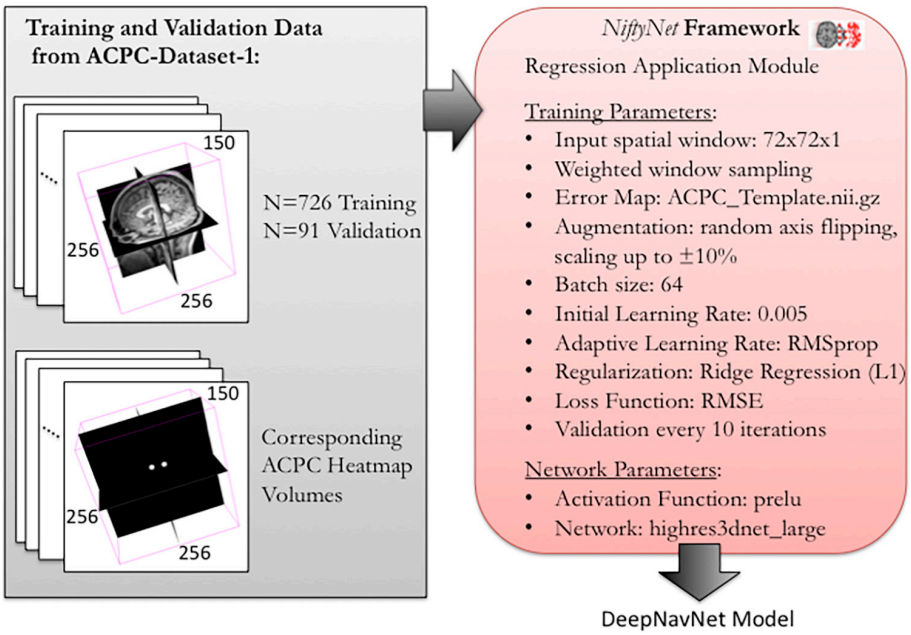


FIGURE 2 | Training and validation of DeepNavNet model using NiftyNet framework. Each input data volume is 256 × 256 × 150 voxels with 1 mm^3 × 1 mm^3 × 1 mm^3 resolution. Key parameters from a customized NiftyNet configuration file are displayed on the right. These parameters were used to create DeepNavNet Model #1 (see **Table 1**).

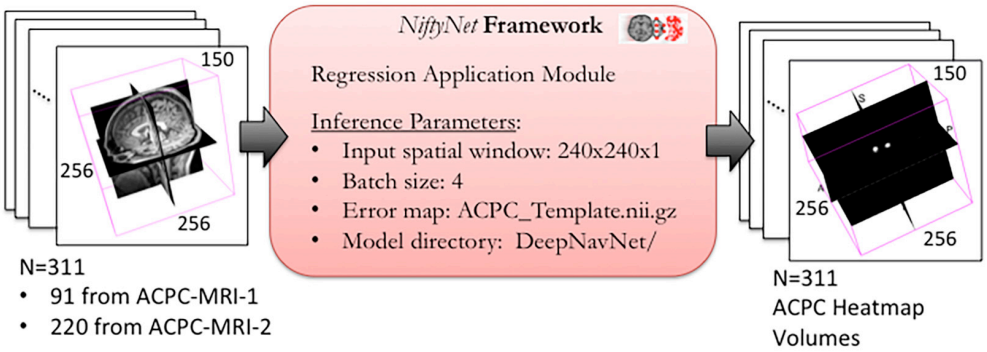


FIGURE 3 | DeepNavNet heatmap regression using NiftyNet Framework for test data.

with outliers kept are presented in **Table 3**. The source of large outliers for the baseline method is unknown and inconsistent; however, visual inspection indicates that it is unable to reliably detect AC and PC points within images with advanced signs of dementia, e.g., atrophy, enlarged ventricles, and asymmetries across the hemispheres. In contrast, the DeepNavNet models produced a fewer number of outlier errors and the magnitude of the errors were much smaller. Visual inspection of the MRI scans does not indicate the cause of the outlier errors, however, the results were consistent across the DeepNavNet models.

DeepNavNet Model-1 was trained using the RMSprop adaptive learning rate method, coupled with RMSE loss function and L1 regression regularization methods, and demonstrated submillimeter accuracy that is significantly better than the baseline method. In addition to further evaluating models against a second independent dataset (i.e., ACPC-MRI-2), the learning curves for training and validation data were examined to assess how well the learned model generalizes to previously unseen data. The pattern of progression of the training and validation learning curves for our overall best performing model DeepNavNet-1 indicates that the learned model does not overfit the training data and is expected to maintain reasonable accuracy across datasets (Lundervold and Lundervold, 2019; see **Figure 6**).

Another instantiation of our model (DeepNavNet Model-2) achieved comparable accuracy to DeepNavNet Model-1 for PC localization; however, even though it also achieved submillimeter accuracy, its AC localization error was greater than Model-1.

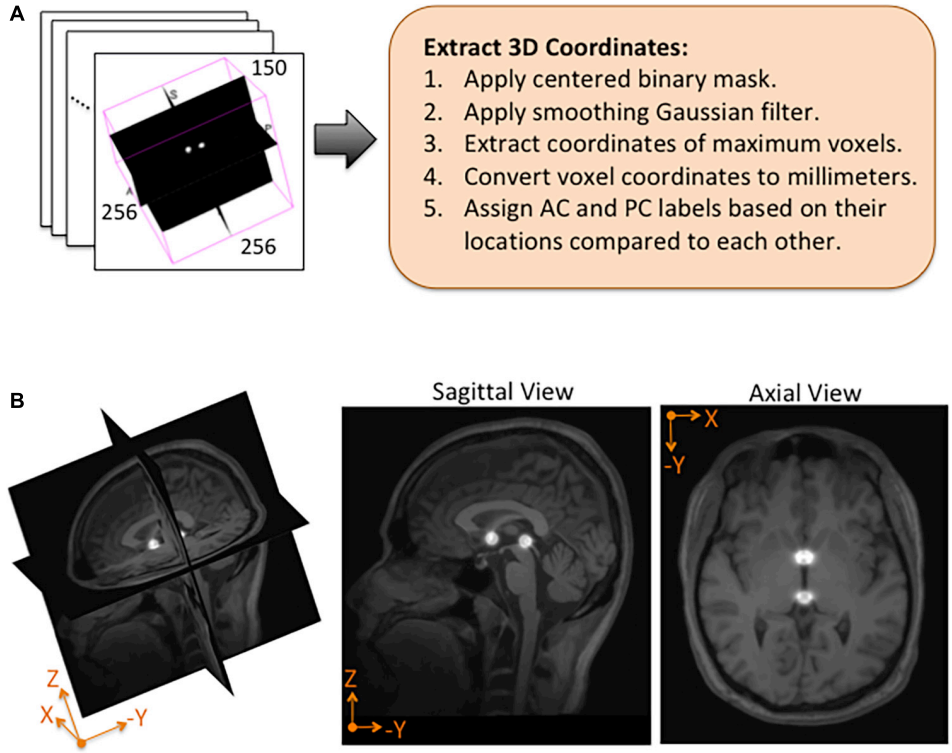


FIGURE 4 | Post-processing and overlay visualizations of a predicted AC-PC heatmap volume. **(A)** The 3D coordinates of AC and PC points are extracted from each predicted heatmap volume. **(B)** Visualizations of a post-processed predicted heatmap overlaid on its corresponding MRI scan. The radius of the spheres is approximately 7.5 mm.

DeepNavNet Model-1 and Model-2 were trained using the same hyper parameter settings, however, the latter used L2 regression regularization instead of L1 regression. The training curves for the DeepNavNet Model-1 and Model-2 follow a very similar stable and gradual progression toward their minimum total loss values across training iterations, with the total loss values for Model-2 being slightly lower than Model-1 (see **Figure 7**).

TABLE 2 | Average 3D AC and PC localization errors of baseline and DeepNavNet models, measured in millimeters.

Model # (Description, N = # test MRIs)	AC Mean ± Std Dev (Max)	PC Mean ± Std Dev (Max)
0 (Baseline)	1.22 ± 0.44 (2.57)	1.00 ± 0.45 (2.31)
1 (RMSE-L1-RMSprop) (*N* = 311)	**0.79 ± 0.33 (1.62)**	**0.78 ± 0.33 (1.66)**
2 (RMSE-L2-RMSprop) (*N* = 311)	0.84 ± 0.35 (1.91)	0.80 ± 0.36 (1.84)
3 (Huber-L1-RMSprop) (*N* = 311)	0.93 ± 0.40 (2.08)	1.00 ± 0.38 (2.14)
4 (Huber-L2-RMSprop) (*N* = 311)	0.81 ± 0.35 (1.85)	0.92 ± 0.39 (2.01)
5 (RMSE-L1-adam) (*N* = 311)	0.87 ± 0.38 (2.05)	0.86 ± 0.35 (1.87)
6 (RMSE-L2-adam) (*N* = 91)	1.01 ± 0.43 (2.08)	1.03 ± 0.42 (2.07)
7 (Huber-L1-adam) (*N* = 91)	X	X
8 (Huber-L2-adam) (*N* = 91)	X	X

An "X" denotes that the Model did not predict heatmaps with AC and PC landmarks highlighted. Models 0–5 were evaluated using the full test set (N = 311), whereas Models 6–8 were only evaluated on the first test set (N = 91) due to their reduced performance. The bolded values correspond to the best performing model as measured by the lowest mean localization errors.

Similarly, DeepNavNet Model-3 and Model-4 were trained using the RMSprop adaptive learning rate technique with L1 and L2 regression regularization techniques, respectively. Although they achieved a significant boost in accuracy compared to baseline (Model-0), they were less accurate than DeepNavNet Model-1 and Model-2. Using the Huber loss function, in contrast to the RMSE function, produced the less accurate DeepNavNet Model-3 and Model-4. Further, their PC localization errors were greater than their AC localizations errors.

DeepNavNet Model-5 significantly outperformed the baseline model and achieved submillimeter accuracy for localizing AC and PC, however, it underperformed compared to DeepNavNet Model-1 and Model-2. The training curve for DeepNavNet Model-5 shows that the model required more training iterations to reach minimum loss values, after which the model then tracks similarly to DeepNavNet, Models 1 and 2 (see **Figure 7**).

DeepNavNet Models 6 to 8 were not evaluated on the ACPC-MRI-2 test set due to the poor quality of their predicted heatmaps from the ACPC-MRI-1 test set. DeepNavNet Model-6 outperformed the baseline for AC localization, however, it was less accurate compared to the other DeepNavNet models. Further,

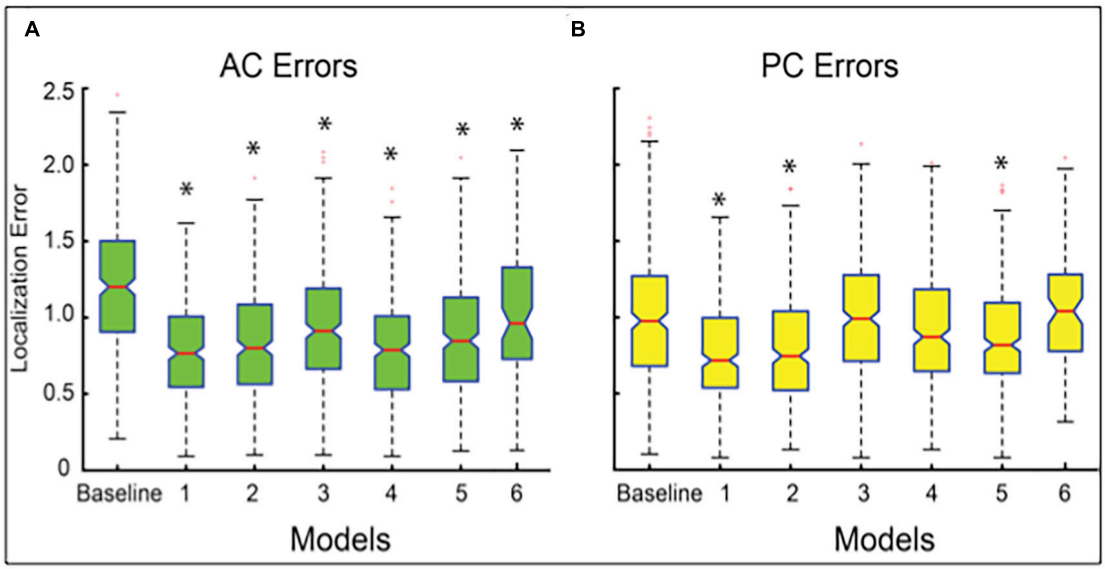


FIGURE 5 | Boxplots of localization errors achieved, measured in millimeters, by the baseline and DeepNavNet models 1-6. Statistical analysis of the baseline and Models 1-5 were conducted using all of the test data (N = 311), whereas Model 6 was evaluated using a subset (N = 91). Box whiskers span 1.5 times the interquartile range (IQR) in both directions. The 95% confidence interval spans 1.57 times the IQR /$\sqrt{n}$ around the median notch for each box. Non-overlapping notches indicate significant differences between models, and asterisks indicate p-values $<$ <0.001 relative to baseline error. **(A)** Boxplots for AC localization errors. **(B)** Boxplots for PC localization errors.

assessment of its total loss across training iterations shows that Model 6 abruptly reached its minimum total loss values sooner, with total loss values larger than high performing instantiations of DeepNavNet models (see **Figure 7**). This is an indication that the initial learning rate (i.e., 0.005) in our configuration may need adjustment. DeepNavNet Models 7 and 8 failed to predict heatmap volumes with the AC and PC points highlighted as indicated by "X" values in **Table 2**. All of the DeepNavNet models that used the Huber loss function (Models 3, 4, 7 and 8), rather than the RMSE loss function, exhibited a lower total loss across training iterations, however, that is not an indication that the model is able to effectively learn to regress to heatmaps with AC and PC points highlighted (see **Figure 7**).

In short, our final resulting DeepNavNet pipeline used our best performing DeepNavNet model (Model-1) that was trained using an RMSE loss function with L1 regression regularization and RMSprop adaptive learning rate techniques. Our resulting DeepNavNet pipeline significantly improved the accuracy of automatically detecting the 3D coordinates for AC and PC points, compared to the baseline method. Intermediate versions of our pipeline that used other versions of our DeepNavNet model had mixed results and overall underperformed compared to our final DeepNavNet model (Model-1).

DISCUSSION

We described the development and evaluation of our novel deep learning-based pipeline that simultaneously detects AC and PC anatomical landmarks with submillimeter accuracy. A core component of this pipeline is the DeepNavNet model which was successfully trained to predict a heatmap with the AC and PC landmarks identified. A key advantage of this method is that one model has been able to detect both AC and PC landmarks within the same predicted heatmap, while preserving the spatial relationship between the landmarks. Furthermore, since all image volumes were normalized to the same template MRI volume with known AC and PC locations, their AC and PC landmarks are expected to be identified within a known region of interest. The DeepNavNet pipeline

TABLE 3 | Number of outliers and the resulting standard deviation and maximum error measured in millimeters with the outliers included for the baseline and our DeepNavNet models.

Model # (Description, N = # test MRIs)	AC # Outliers, Std Dev (Max)	PC # Outliers, Std Dev (Max)
0 (Baseline)	11, 5.01 (62.59)	10, 3.45 (49.76)
1 (RMSE-L1-RMSprop) (N = 311)	**3, 0.38 (3.25)**	**6, 0.36 (2.00)**
2 (RMSE-L2-RMSprop) (N = 311)	2, 0.38 (2.94)	1, 0.36 (1.91)
3 (Huber-L1-RMSprop) (N = 311)	4, 0.42 (2.38)	0, 0.38 (2.14)
4 (Huber-L2-RMSprop) (N = 311)	4, 0.39 (2.31)	3, 0.41 (2.74)
5 (RMSE-L1-adam) (N = 311)	1, 0.39 (2.49)	4, 0.37 (2.30)
6 (RMSE-L2-adam) (N = 91)	6, 11.39 (79.40)	4, 8.09 (72.65)

The bolded values correspond to the best performing model as measured by the lowest mean localization errors.



FIGURE 6 | Training and validation learning curves for DeepNavNet Model-1.

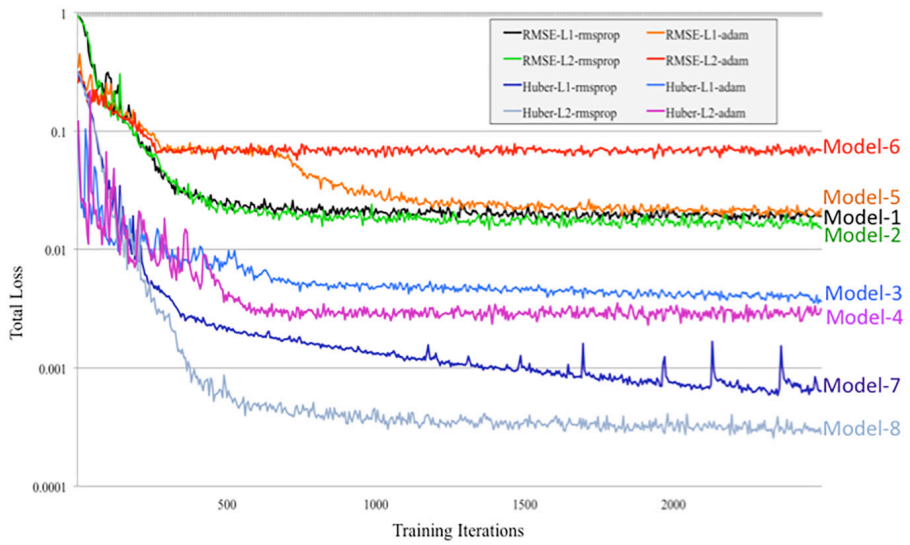


FIGURE 7 | Tensorboard training learning curves for DeepNavNet Models 1-8 using hyperparameters described in **Table 1**.

demonstrated that it is able to automatically identify and fine-tune predicted AC and PC locations with submillimeter accuracy. Post-processing to extract the 3D coordinates for each landmark within the heatmap volumes was relatively simple since the regions of interest were confined to regions around the initial AC and PC template locations. More sophisticated methods to extract centroid coordinates of other anatomical structures from volumetric heatmaps could be considered in future applications, such as a DeepNavNet pipeline to automatically locate specific deep brain stimulation targets.

A key component of our research included the creation of a large-scale annotated dataset of MRI volumes, which were critical for training, validating and testing our models, and will be invaluable for follow-up research within the neuroimaging community. We manually annotated the AC and PC landmark locations for 1,128 1.5-T MRI volumes collated from publicly available data collected from multiple institutions which were acquired from healthy subjects to those with varying degrees of cognitive decline. The anatomical brain structures of subjects with advanced stages of dementia were visibly different from relatively healthy subjects, e.g., enlarged ventricles, atrophy, and asymmetries. As such, manual identification of the AC and PC points took longer, while also providing a more diverse dataset for training and evaluating more robust models. Even so, it is expected that the DeepNavNet models and pipeline would require fine-tuning for different types of MRI scans and other neuroimaging modalities.

The focus of this work was to develop a deep learning based pipeline that demonstrates consistent submillimeter accuracy for detecting AC and PC points, across a large-scale diverse dataset of 1.5-T MRI volumes, as a first step toward an automation pipeline to enable a modernized human-machine teaming approach to

surgery planning. Further improvements to accuracy may be achieved by experimenting with other loss functions, such as an adaptive loss function designed for landmark localization (Teixeira et al., 2019). Although we achieved superior accuracy results, the practical implementation of this work will require further research to reduce the overall computation time. In particular, the heatmap inference time per MRI scan is a few minutes, whereas the post-processing step to extract the coordinates is approximately a second per MRI scan. We experimented with multiple combinations of different training parameters, however, further research could leverage techniques that learn optimal hyperparameters (Liaw et al., 2018). Also, meta-learning approaches to identify optimal parameters may reduce the size of the resulting models and enable a more efficient network for inference (Vanschoren, 2018). Furthermore, with the rise of edge computing, there is growing interest in improving the efficiency of deep neural networks to enable deployment of such technologies on edge devices. Methods to prune deep neural networks are expected to enable more efficient models (Blalock et al., 2020). In addition, in April 2020, the creators of NiftyNet framework announced that they are shifting toward an open-source PyTorch-based deep learning framework for medical imaging called Medical Open Network for AI (MONAI).[2] Future experiments to advance DeepNavNet would likely benefit from moving to the MONAI framework as well since it is expected to gain momentum in the deep learning medical imaging community. Another machine learning approach that is gaining momentum across multiple application domains, including landmark detection in medical images, combines deep learning and reinforcement learning methods – i.e., deep reinforcement learning (DRL) (Mnih et al., 2015; Francois-Lavet et al., 2018; Sutton and Barto, 2018). In contrast to supervised deep learning approaches, which learn to minimize a loss function based on labeled training data, a DRL artificial agent learns to maximize an award function by trial-and-error as it navigates the data environment in a sequential Markov decision process (MDP) framework. Ghesu et al. (2016) introduced an approach that used a DRL agent to locate anatomical landmarks in medical image volumes. Building upon this research, DRL methods continue to advance and show promise as an effective strategy for real-time navigation to landmarks within medical images (Ghesu et al., 2017, 2019; Al and Yun, 2019; Alansary et al., 2019). Even so, establishing an effective decision space and award function that reliably converges is non-trivial for DRL approaches in dynamic environments, and is especially a barrier for using such agents where consistently accurate and safe results are required (Dulac-Arnold et al., 2019). A multi-agent collaborative DRL approach, introduced in 2019, achieved an average localization error of 0.93 ± 0.18 and 1.05 ± 0.25 mm for AC and PC points, respectively (Vlontzos et al., 2019). Although our current supervised deep learning approach is more accurate, there is an acceleration of innovation in DRL methods and such artificial agents are expected to improve through experience as they navigate through more neuroimaging data, such as our ACPC-MRI -1 and ACPC-MRI-2 datasets.

[2]Medical Open Network for Artificial Intelligence available from https://monai.io

Our method described here focused on identifying AC and PC points as a necessary first step toward aligning to a stereotactic atlas. However, the regression pipeline that we have developed and validated can be applied toward segmentation of a wide range of neural and non-neural structures. The heatmap is a relatively straightforward approach to generate probabilistic distributions of target region locations, enabling rapid model development and training. Post-processing algorithms for region coordinate extraction can then employ more advanced segmentation methods, such as spherical Hough transforms or an additional convolutional neural network. Our model presented herein provides a framework upon which future parcellation networks can be developed. Additional exploration will be needed to optimize our method for other biological structures (e.g., hyperparameter tuning, model architecture, etc.).

Another approach to consider beyond improved indirect targeting using AC and PC landmarks is to employ deep learning methods to directly register a subject's MRI data to a stereotactic atlas, followed by an indirect targeting process (Sutton and Barto, 2018). In addition, ultra high-field 7.0 T MRI scanners are becoming clinically available at a limited number of large medical institutions (e.g., Mayo Clinic, Rochester, MN, United States), and provide unprecedented high-resolution imaging of deep brain structures (Cao et al., 2018). As such, this technology is expected to enable direct targeting for deep brain stimulation and other therapies. Furthermore, they will provide invaluable data to improve and enable AI-powered surgical planning systems that would learn to safely navigate to deep brain stimulation targets for optimal clinical results. We anticipate that our model will benefit from the incorporation of ultra-high field imaging via direct localization of target nuclei and pathology.

CONCLUSION

Our novel DeepNavNet pipeline automatically identifies the 3D coordinates of AC and PC anatomical landmarks in MRI volumetric data with state-of-the-art submillimeter accuracy, as required for image-guided targeted neurosurgery planning procedures. This technology is a first step toward automating surgical planning to improve surgical accuracy and efficiency.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.oasis-brains.org and https://mindboggle.info/data.html.

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.

AUTHOR CONTRIBUTIONS

CE designed the methodology and executed the experiments, performed data collation and annotation, developed the DeepNavNet software pipeline, conducted analyses of experimental results, and drafted and revised the manuscript. AG annotated data, conducted statistical analyses, and edited the manuscript. AR annotated data and edited the manuscript. AK provided guidance and edited the manuscript. KL annotated data, provided guidance, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by The Grainger Foundation.

ACKNOWLEDGMENTS

We thank Joshua Jacobs who collated the ACPC-MRI-1 dataset, established our annotation pipeline and annotated a subset of data; and Benjamin Gifford and Kristen H. Lee for their annotation support.

REFERENCES

Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., and Erickson, B. J. (2017). Deep learning for brain MRI segmentation: state of the art and future directions. *J. Digit. Imaging* 30, 449–459. doi: 10.1007/s10278-017-9983-4

Al, W. A., and Yun, I. D. (2019). Partial policy-based reinforcement learning for anatomical landmark localization in 3D medical images. *IEEE Trans. Med. Imaging* 39, 1245–1255. doi: 10.1109/TMI.2019.2946345

Alansary, A., Oktay, O., Li, Y., Folgoc, L. L., Hou, B., Vaillant, G., et al. (2019). Evaluating reinforcement learning agents for anatomical landmark detection. *Med. Image Anal.* 53, 156–164. doi: 10.1016/j.media.2019.02.007

Ardekani, B. A., and Bachman, A. H. (2009). Model-based automatic detection of the anterior and posterior commissures on MRI scans. *Neuroimage* 46, 677–682. doi: 10.1016/j.neuroimage.2009.02.030

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025

Berger, L., Hyde, E., Cardoso, M. J., and Ourselin, S. (2017). An adaptive sampling scheme to efficiently train fully convolutional networks for semantic segmentation. *arXiv*[Preprint] arXiv 1709.02764,

Bhanu Prakash, K. N., Hu, Q., Aziz, A., and Nowinski, W. L. (2006). Rapid and automatic localization of the anterior and posterior commissure point landmarks in MR volumetric neuroimages. *Acad. Radiol.* 13, 36–54. doi: 10.1016/j.acra.2005.08.023

Binder, D. K., Rau, G. M., and Starr, P. A. (2005). Risk factors for hemorrhage during microelectrode-guided deep brain stimulator implantation for movement disorders. *Neurosurgery* 56, 722–732;discussion722–732. doi: 10.1227/01.neu.0000156473.57196.7e

Blalock, D., Gonzalez Ortez, J. J., Frankle, J., and Guttag, J. (2020). "What is the state of neural network pruning," in *Machine Learning and Systems 2020 (MLSys 2020)*. Austin, TX.

Cao, X., Yang, J., Wang, L., Wang, Q., and Shen, D. (2018). "Non-ridgid brain MRI registration using two-stage deep perceptive networks", in: *Joint Annual Meeting ISMRM-ESMRMB*. 1176. Paris, France.

Cho, Z. H., Min, H. K., Oh, S. H., Han, J. Y., Park, C. W., Chi, J. G., et al. (2010). Direct visualization of deep brain stimulation targets in Parkinson disease with the use of 7-tesla magnetic resonance imaging. *J. Neurosurg.* 113, 639–647. doi: 10.3171/2010.3.JNS091385

D'Haese, P. F., Cetinkaya, E., Konrad, P. E., Kao, C., and Dawant, B. M. (2005). Computer-aided placement of deep brain stimulators: from planning to intraoperative guidance. *IEEE Trans. Med. Imaging* 24, 1469–1478. doi: 10.1109/TMI.2005.856752

Dulac-Arnold, G., Mankowitz, D., and Hester, T. (2019). "Challenges of real-world reinforcement learning," in *Proceedings of the International Conference on Machine Learning (ICML)*, (Long Beach, CA).

Edwards, C. A., Kouzani, A., Lee, K. H., and Ross, E. K. (2017). Neurostimulation devices for the treatment of neurologic disorders. *Mayo Clin. Proc.* 92, 1427–1444. doi: 10.1016/j.mayocp.2017.05.005

Edwards, C. A., Rusheen, A. E., Oh, Y., Paek, S. B., Jacobs, J., Lee, K. H., et al. (2018). A novel re-attachable stereotactic frame for MRI-guided neuronavigation and its validation in a large animal and human cadaver model. *J. Neural Eng.* 15:066003. doi: 10.1088/1741-2552/aadb49

Elias, W. J., Lipsman, N., Ondo, W. G., Ghanouni, P., Kim, Y. G., Lee, W., et al. (2016). A randomized trial of focused ultrasound thalamotomy for essential tremor. *N. Engl. J. Med.* 375, 730–739. doi: 10.1056/NEJMoa1600159

Falk, T., Mai, D., Bensch, R., Cicek, O., Abdulkadir, A., Marrakchi, Y., et al. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16, 67–70. doi: 10.1038/s41592-018-0261-2

Fedorov, A., Johnson, J., Damaraju, E., Ozerin, A., Calhoun, V. D., and Plis, S. (2016). End-to-end learning of brain tissue segmentation from imperfect labeling. *arXiv*[Preprint] arXiv 1612.00940,

Francois-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., and Pineau, J. (2018). *An Introduction to Deep Reinforcement Learning*. Boston MA: NOW.

Ghesu, F. C., Georgescu, B., and Comaniciu, D. (2017). "Robust multi-scale anatomical landmark detection in incomplete 3D-CT data," in *Proceedings of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2017)*, eds M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. Collins, and S. Duchesne (Cham: Springer).

Ghesu, F. C., Georgescu, B., Mansi, T., Neumann, D., Hornegger, J., and Comaniciu, D. (2016). "An artificial agent for anatomical landmark detection in medical images," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*, eds S. Ourselin, L. Joskowicz, M. Sabuncu, G. Unal, and W. Wells (Cham: Springer), 229–237.

Ghesu, F. C., Georgescu, B., Zheng, Y., Grbic, S., Maier, A., Hornegger, J., et al. (2019). Multi-scale deep reinforcement learning for real-time 3d-landmark detection in CT scans. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 176–189. doi: 10.1109/TPAMI.2017.2782687

Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D. I., Wang, G., et al. (2018). NiftyNet: a deep-learning platform for medical imaging. *Comput. Methods Programs. Biomed.* 158, 113–122. doi: 10.1016/j.cmpb.2018.01.025

Grewal, S. S., Middlebrooks, E. H., Kaufmann, T. J., Stead, M., Lundstrom, B. N., Worrell, G. A., et al. (2018). Fast gray matter acquisition T1 inversion recovery MRI to delineate the mammillothalamic tract for preoperative direct targeting of the anterior nucleus of the thalamus for deep brain stimulation in epilepsy. *Neurosurg. Focus* 45:E6. doi: 10.3171/2018.4.FOCUS18147

Grewal, S. S., Middlebrooks, E. H., Okromelidze, L., Gosden, G. P., Tatum, W. O., Lundstrom, B. N., et al. (2020). Variability between direct and indirect targeting of the anterior nucleus of the thalamus. *World Neurosurg.* 139, e70–e77. doi: 10.1016/j.wneu.2020.03.107

Hariz, M. I., and Fodstad, H. (1999). Do microelectrode techniques increase accuracy or decrease risks in pallidotomy and deep brain stimulation? A critical review of the literature. *Stereotact. Funct. Neurosurg.* 72, 157–169. doi: 10.1159/000029720

Hartmann, C. J., Fliegen, S., Groiss, S. J., Wojtecki, L., and Schnitzler, A. (2019). An update on best practice of deep brain stimulation in Parkinson's disease. *Ther. Adv. Neurol. Disord* 12:1756286419838096. doi: 10.1177/1756286419838096

Hinton, G. (2018). Deep learning-a technology with the potential to transform health care. *JAMA* 320, 1101–1102. doi: 10.1001/jama.2018.11100

Isaacs, B. R., Keuken, M. C., Alkemade, A., Temel, Y., Bazin, P. L., and Forstmann, B. U. (2020). Methodological considerations for neuroimaging in deep brain stimulation of the subthalamic nucleus in Parkinson's disease patients. *J. Clin. Med.* 9:3124. doi: 10.3390/jcm9103124

Kikinis, R., Pieper, S. D., and Vosburgh, K. G. (2014). "3D slicer: a platform for subject-specific image analysis, visualization, and clinical support," in

Proceedings of the Intraoperative Imaging and Image-Guided Therapy, ed. F. A. Jolesz (New York, NY: Springer).

Kingma, D. P., and Ba, J. L. (2015). "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, (San Diego, CA).

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M. C., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46, 786–802. doi: 10.1016/j.neuroimage.2008.12.037

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Hame, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLoS Comput. Biol.* 13:e1005350. doi: 10.1371/journal.pcbi.1005350

Klein, A., and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* 6:171. doi: 10.3389/fnins.2012.00171

Korfiatis, P., and Erickson, B. (2019). Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas. *Clin. Radiol.* 74, 367–373. doi: 10.1016/j.crad.2019.01.028

LaMontagne, P. J., Benzinger, T. L. S., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., et al. (2019). OASIS-3: longitudinal neuroimaging, clinical, and cognitive datatset for normal aging and alzheimere disease. *medRxiv* doi: 10.1101/2019.12.13.19014902

Lathuiliere, S., Mesejo, P., Alameda-Pineda, X., and Horaud, R. (2019). A comprehensive analysis of deep regression. *IEEE Trans. Pattern Anal. Mach. Intell* 42, 2065–2081. doi: 10.1109/TPAMI.2019.2910523

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

Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M. J., and Vercauteren, T. (2017). "On the compactness, efficience, and representation of 3d convolutional networks: brain parcellation as a pretext task," in *Proceedings of the Information Processing in Medical Imaging. IPMI 2017. Lecture Notes in Computer Science*, ed. M. Niethammer (London: University College).

Li, Z., Zhang, J. G., Ye, Y., and Li, X. (2016). Review on factors affecting targeting accuracy of deep brain stimulation electrode implantation between 2001 and 2015. *Stereotact. Funct. Neurosurg.* 94, 351–362. doi: 10.1159/000449206

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., and Stoica, I. (2018). "Tune: a research platform for distributed model selection and training," in *Proccedings of the 2018 International Conference on Machine Learning AutoML Workshop*.

Liu, M., Zhang, J., Adeli, E., and Shen, D. (2018). Landmark-based deep multi-instance learning for brain disease diagnosis. *Med. Image Anal.* 43, 157–168. doi: 10.1016/j.media.2017.10.005

Liu, Y., and Dawant, B. M. (2015). Automatic localization of the anterior commissure, posterior commissure, and midsagittal plane in MRI scans using regression forests. *IEEE J. Biomed. Health Inform.* 19, 1362–1374. doi: 10.1109/JBHI.2015.2428672

Lundervold, A. S., and Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* 29, 102–127. doi: 10.1016/j.zemedi.2018.11.002

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* 19, 1498–1507. doi: 10.1162/jocn.2007.19.9.1498

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi: 10.1038/nature14236

Pallavaram, S., Dawant, B. M., Koyama, T., Yu, H., Neimat, J., Konrad, P. E., et al. (2009). Validation of a fully automatic method for the routine selection of the anterior and posterior commissures in magnetic resonance images. *Stereotact. Funct. Neurosurg.* 87, 148–154. doi: 10.1159/000209295

Pallavaram, S., Yu, H., Spooner, J., D'Haese, P. F., Bodenheimer, B., Konrad, P. E., et al. (2008). Intersurgeon variability in the selection of anterior and posterior commissures and its potential effects on target localization. *Stereotact. Funct. Neurosurg.* 86, 113–119. doi: 10.1159/000116215

Panesar, S. S., Kliot, M., Parrish, R., Fernandez-Miranda, J., Cagle, Y., and Britz, G. W. (2020). Promises and Perils of Artificial Intelligence in Neurosurgery. *Neurosurgery* 87, 33–44. doi: 10.1093/neuros/nyz471

Payer, C., Stern, D., Bischof, H., and Urschler, M. (2016). "Regressing heatmaps for multiple landmark localization using CNNs," in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, eds S. Ourselin, L. Joskowicz, M. Sabuncu, G. Unal, and W. Wells (Cham: Springer).

Payer, C., Stern, D., Bischof, H., and Urschler, M. (2019). Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Med. Image Anal.* 54, 207–219. doi: 10.1016/j.media.2019.03.007

Ronneberger, O., Firscher, P., and Brox, T. (2015). "U-Net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, eds N. Navab, J. Hornegger, W. Wells, and A. Frangi (Cham: Springer), 234–241.

Slavin, K. V., Thulborn, K. R., Wess, C., and Nersesyan, H. (2006). Direct visualization of the human subthalamic nucleus with 3T MR imaging. *AJNR Am. J. Neuroradiol.* 27, 80–84.

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge MA: The MIT Press.

Teixeira, B., Tamersoy, B., Singh, V., and Kapoor, A. (2019). Adaloss: adaptive loss function for landmark localization. *arXiv*[Preprint] arXiv:1908.01070,

Vanschoren, J. (2018). Meta-learning: a survey. *arXiv*[Preprint] arXiv 1810.03548,

Verard, L., Allain, P., Travere, J. M., Baron, J. C., and Bloyet, D. (1997). Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis. *IEEE Trans. Med. Imaging* 16, 610–616. doi: 10.1109/42.640751

Vlontzos, A., Alansary, A., Kamnitsas, K., Rueckert, D., and Kainz, B. (2019). "Multiple landmark detection using multi-agent reinforcement learning," in *Proceedings of the 29th International Conference on Medical Imaging Computing and Computer Assisted Intervention - MICCAI 2019*, ed. D. Shen (Cham: Springer).

Zhang, J., Liu, M., and Shen, D. (2017). Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. *IEEE Trans. Image Process.* 26, 4753–4764. doi: 10.1109/TIP.2017.2721106

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 23 June 2021
doi: 10.3389/fnins.2021.690044

Possibilistic Clustering-Promoting Semi-Supervised Learning for EEG-Based Emotion Recognition

Yufang Dan[1†], Jianwen Tao[1†], Jianjing Fu[2] and Di Zhou[3]*

[1] Institute of Artificial Intelligence Application, Ningbo Polytechnic, Ningbo, China, [2] School of Media Engineering, Communication University of Zhejiang, Hangzhou, China, [3] Dazhou Industrial Technological Institute of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou, China

The purpose of the latest brain computer interface is to perform accurate emotion recognition through the customization of their recognizers to each subject. In the field of machine learning, graph-based semi-supervised learning (GSSL) has attracted more and more attention due to its intuitive and good learning performance for emotion recognition. However, the existing GSSL methods are sensitive or not robust enough to noise or outlier electroencephalogram (EEG)-based data since each individual subject may present noise or outlier EEG patterns in the same scenario. To address the problem, in this paper, we invent a Possibilistic Clustering-Promoting semi-supervised learning method for EEG-based Emotion Recognition. Specifically, it constrains each instance to have the same label membership value with its local weighted mean to improve the reliability of the recognition method. In addition, a regularization term about fuzzy entropy is introduced into the objective function, and the generalization ability of membership function is enhanced by increasing the amount of sample discrimination information, which improves the robustness of the method to noise and the outlier. A large number of experimental results on the three real datasets (i.e., DEAP, SEED, and SEED-IV) show that the proposed method improves the reliability and robustness of the EEG-based emotion recognition.

Keywords: semi-supervised classification, membership function, electroencephalogram, emotion recognition, fuzzy entropy

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Yue Zhao, Harbin Institute of Technology, China
Jin Cui, Northwest University, China

***Correspondence:**
Jianjing Fu
fjjmsn@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *02 April 2021*
Accepted: *28 April 2021*
Published: *23 June 2021*

Citation:
Dan Y, Tao J, Fu J and Zhou D (2021) Possibilistic Clustering-Promoting Semi-Supervised Learning for EEG-Based Emotion Recognition. Front. Neurosci. 15:690044. doi: 10.3389/fnins.2021.690044

INTRODUCTION

Emotion is embodied by human beings: we are born with an innate understanding of emotion (Dolan, 2002; Zhang et al., 2016, 2019b). The complexity of emotion leads to different people's understanding of emotion. Therefore, it is more difficult for machines to accurately understand emotion. As one of the hottest research topics in the field of affective computing, emotion recognition has received extensive attention from the field of pattern recognition and brain neural research (Kim et al., 2013; Mühl et al., 2014; Chu et al., 2017). In this work, we focus on emotional speculation through changes in the body. Basically, the representative internal changes of the body include blood pressure, magneto encephalogram, electroencephalogram (EEG), heart rate, respiratory rate (Mühl et al., 2014), and so on. The EEG-based traditional emotion recognition system usually has two parts: feature extraction and recognizer training (Lan et al., 2019;

Zhang et al., 2020b). Jenke et al. (2014) made a comprehensive review on EEG feature extraction methods. In order to solve the recognition problem, many EEG-based emotion recognition methods have been provided recently (Musha et al., 1997; Kim et al., 2013). An ideal emotion-based brain–computer interface (BCI) can detect the emotional state through spontaneous EEG signals without explicit input from the user (Zhang et al., 2019b) and make a corresponding response to different emotional state. This kind of BCI may enhance the consumer experience in the time of an interactive session. Therefore, different approaches (Zhang et al., 2016, 2017) have been designed to recognize various emotion signals from brain waves. The latest affective BCIs (aBCI) have taken machine learning algorithms and depend on a few features with discriminative information (Jenke et al., 2014; Mühl et al., 2014). A representation of how aBCI exemplification operates is described here. When recording EEG signals, in order to generate a desired target emotion signal, it is necessary to provide users with affective stimulation of specific emotions. In the training/calibration session, the required features and corresponding emotion labels are extracted from EEG signals to train the classifier. In an ongoing BCI session, the feature extractor receives the real-time EEG data, sending the extracted features to the classifier for real-time affection classification. In this paradigm (Mühl et al., 2014), many researchers have reported a pleasing classification performance.

While effective machine learning and deep learning require a large amount of labeled data, sufficient labeled data are often difficult to obtain in real applications. Although manually labeled instances can make up for the lack of labeled instance to a certain extent, this process is time-consuming and laborious. Then, the semi-supervised learning (SSL; Zhou et al., 2003, 2014; Zhu et al., 2003; Chapelle et al., 2006; Zhu, 2008; Zhu and Goldberg, 2009; Gao et al., 2010; Zhou and Li, 2010; Zhao and Zhou, 2018; Tao et al., 2019; Wang Q.-W. et al., 2019; Wang T.-Z. et al., 2019; Zhang et al., 2019c) technique was proposed, which learns a model from a small amount of labeled instances and a large amount of unlabeled instances and solves the problem of insufficient labeled instance (i.e., poor generalization of the model obtained by supervised learning and inaccurate models obtained by unsupervised learning). Tu and Sun (2013) proposed a semi-supervised feature extraction method for EEG classification. Wu and Deng (2018) and Tao et al. (2015, 2016, 2017) proposed a semi-supervised classification framework based on collaborative training and differential evolution to improve the impact of random initial values of input layer parameters of neural networks on classification. Zu et al. (2019) explored to invent a semi-supervised classification method for large-scale remote sensing images based on low-rank block maps, and the reseults have been used to effectively improve classification performance, as graph-based semi-supervised learning (GSSL; Li and Zhou, 2011; Liu et al., 2012; Wang et al., 2012), with its intuitiveness and good learning performance, has been extensively studied. GSSL has two different types of inference, namely, transductive inference (Zhou et al., 2003; Zhu et al., 2003; Wang and Zhang, 2008; Wang et al., 2017) and inductive inference (Belkin et al., 2006; Gao et al., 2010; Nie et al., 2010). The transductive inference assumes that the unlabeled data in the learning process is exactly the test data, and it does not have a good prediction effect on the out-of-sample, for example, LGC (Zhou et al., 2003), GFHF (Zhu et al., 2003), LNP (Wang and Zhang, 2008), and ACA-S3VM (Wang et al., 2017), etc. The inductive inference puts all the instances together in the assumption learning process to find their commonalities and then gets a model. The important point to note is that the test instance does not exist in the training dataset. The Manifold Regularization (MR; Belkin et al., 2006) is a very common inductive GSSL inference, such as, GLSSVM (Gao et al., 2010), FME/U (Nie et al., 2010), etc., and the FME/U was proposed by Nie et al. (2010) generalized the MR framework induction.

Generally, GSSL inference requires certain assumptions. While the GSSL inference models have been employed on EEG datasets due to its effectiveness and intuitiveness, limited effort has been made on improving its performance by the clustering assumption. One of the most common assumptions is the clustering hypothesis: "Similar instances should share the same class label" (Chapelle et al., 2006; Zhu and Goldberg, 2009; Xue et al., 2011; Zhou et al., 2014; Wang Q.-W. et al., 2019; Wang T.-Z. et al., 2019). The assumption has an implicit assumption each instance should clearly belong to a certain class. We call this kind of classification hard classification. However, in real emotion recognition applications, it is difficult to strictly employ this assumption. For example, the same emotion will be understood as different emotion recognition by different subject at different/same scenario.

In order to solve the hard classification problem based on the traditional clustering assumption, Wang et al. (2012) and Zhang et al. (2019a) proposed a new semi-supervised classification method based on modified cluster assumption (SSCCM), which is a soft classification method based on clustering assumption. It constrained the similar instances that share the same label membership. Each instance could belong to multiple class labels and have corresponding membership values, which made good use of the fuzzy clustering assumption (Krishnapuram and Keller, 1993). However, its constraint condition made the total membership of each instance for different labels be 1, which may cause the label membership of some noises to be the same as the label membership of some normal instance, even for one or more classes. The label membership value of the noise may be greater than the normal instance, that is, the correlation is greater, which will cause misrecognition due to its constraint.

Toward the problem of the SSCCM method, we further develop a Possibilistic Clustering Promoting semi supervised learning for EEG-based Emotion Recognition (PCP-ER). The main idea of the method is threefold. First, each instance and its local weighted mean LWM (Bottou and Vapnik, 1992; Atkeson et al., 1997; Xue and Chen, 2007) share the similar memberships. Then, the recognition results obtained by the decision function and membership function are used to verify each other for enhancing the reliability of semi-supervised classification learning method. Finally, a regularization term about fuzzy entropy is added to increase the amount of sample discrimination information. We then obtain a membership function with stronger generalization ability, thereby overcoming

the interference of noise and outlier on the recognition result, and further improving the robustness of the recognition method. In sum, the main contributions of this paper as follows:

(1) A possibilistic clustering promoting semi supervised learning for EEG-based emotion recognition (called as PCP-ER shortly) is proposed;
(2) This method introduces a regularization term about fuzzy entropy to obtain a label membership function with more generalization to overcome the influence of noise and outlier and improve the robustness of the method;
(3) A serial of experiments performed on real-world EEG datasets (i.e., DEAP, SEED, and SEED-IV) to verify the robust effectiveness and recognition reliability of the proposed framework.

The rest of paper is organized as follows. We design our framework PCP-ER in section "Proposed Framework" followed by its corresponding optimal algorithm in section "Optimization." The Algorithm is explained in section "Algorithm description." Section "Discussion" gives algorithm analysis including the reliability, convergence and generalization error bound. Experimental results and analysis on three real-world EEG datasets (i.e., DEAP, SEED, and SEED-IV) are reported in section "Experiment." Finally, we draw a conclusion in section "Conclusion."

PROPOSED FRAMEWORK

Problem Statement

In real classification applications, there are some examples of EEG-based semi-supervised clustering methods where it is difficult to assign an instance explicitly belongs to only one class, such as those boundary instances. Since the hard clustering assumption implicitly constrained each instance that has a clear label assignment, the distribution of real data cannot fully be reflected, and the distribution of these boundary instances may be changed. Therefore, when a semi-supervised classification method adopts this assumption, the predictions on those boundary instances are not good. Wang et al. (2012) and Zhang et al. (2020a) proposed the classification method with modified clustering assumption to a certain extent improved the performance of the classification method based on the hard clustering assumption. Each instance will have a label membership value of a different class rather than only belonging to one class. It can reduce the "misleading" classification impact of those boundary instances. **Figure 1** gives an example, in which the data in both Class1 and Class2 with a question mark are unlabeled data, the rest of other instances are labeled data, and the dashed line is the middle dividing line of the two classes. $x1$ can be regarded as a boundary point or an outlier point. $x2$ is certainly more like an instance of Class1 and class2 than $x1$. However, following the SSCCM, the instance is closer to one class, the membership value about this class is larger and *vice versa*. Therefore, the membership values of $x2$ belonging to class1 and class2 are 0.5 and 0.5, respectively. The membership values of instance $x1$ belonging to class1 and class2 are 0.6 and 0.4, respectively. The membership values of $x1$ belonging to class1, which is larger than that of $x2$, making $x1$ more likely to be a normal instance and $x2$ an outlier. The mainly reason is the constraint term that the sum of membership from different classes of a single instance is always 1 in SSCCM, even if it is a boundary point or an outlier such as $x1$.

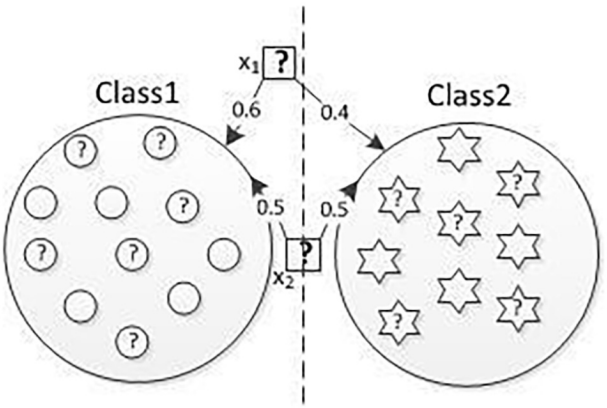


FIGURE 1 | Problem description.

In order to overcome the influence of noise and outlier data on classifiers, we propose a PCP-ER.

Formulation

In order to ensure the classification method of clustering assumption has better classification reliability and robustness, the PCP-ER method achieves the following three goals: (1) any instance should have similar label membership to its corresponding LWM; (2) the decision function and the membership function can mutually verify the classification results of a test instance and have convergence; and (3) we should try to overcome the influence caused by noise and outlier. The classification method proposed in this paper will calculate the LWM of each instance by Euclidean distance and will then obtain the decision function and label membership function through the objective function based on the square loss function with an alternating iterative strategy and utilize the fuzzy entropy to overcome the influence of noise and outlier, thereby improving the robustness of the method. Finally, an optimized classifier model with double verification is constructed by a decision function $f(x)$ and a membership function $w(x)$.

Let dataset $X = \{x_1, x_2, ..., x_i, x_{i+1}, ..., x_n\}$, where $X_l = \{x_i\}_{i=1}^{l}$ is the labeled data with the corresponding labels $Y_l = \{y_1, y_2, ..., y_l\}^T \in \mathbb{R}^{l\times M}$, and n is the total number of instances, $l \ll n$. $X_u = \{x_j\}_{j=l+1}^{n}$ is the unlabeled data, where $x_i \in R^d$ is the $i-th$ instance with d dimensions. The LWM of each x_i (i.e., $\widehat{x}$) is defined as

$$\hat{x}_i = \frac{\sum_{x_j \in Ne(x_i)} W_{ij} x_j}{\sum_{x_j \in Ne(x_i)} W_{ij}}, \tag{1}$$

Here, $Ne(x_i)$ is composed by k nearest neighbors of x_i, each of them is measured by the Euclidean Distance. $G = (X, W)$

denotes a undirected weight graph, where $W \in \mathbb{R}^{n\times n}$ is a weight matrix, $W_{ji} = W_{ij} \geq 0$, and the element of W is measured by

$$W_{ij} = \begin{cases} \exp(-\gamma\|x_i - x_j\|^2), & x_i \text{ is the nearest neighbor of } x_j \\ 0 & \text{otherwise} \end{cases},$$

where γ controls the local scope of the Gaussian kernel function. The larger γ is, the smaller the local scope (i.e., the width) is and *vice versa*. When γ is fixed, W_{ij} decreases monotonically with the increase in the distance between x_i and x_j. Therefore, the clustering problem is transformed into a graph problem. $\widehat{X}_l = \{\widehat{x}\}_{i=1}^{l}$ and $\widehat{X}_u = \{\widehat{x}\}_{i=l+1}^{n}$ are the LWM of l labeled data and $(n-l)$ unlabeled data, respectively. $\{c_m\}_{m=1}^{M}$ is the coded representation of Mclasses. If x_i belongs to the $m-$th class, then $y_i = c_m$, the label and the category encoding are encoded according to one of the M categories, that is, the both of the label and the category coding is a vector of dimension M so PCP-ER can be directly applied to multi-class classification tasks. Let $y_i \in \mathbb{R}^{1\times M}$ and $c_m \in \mathbb{R}^{1\times M}$. If x_i belongs to the $m-$th class, then the $m-$th element of y_i is designated as 1, that is, $y_{im} = 1, m = 1, 2, ..., M$, and the other elements of y_i are 0. $y_{io} = 0, o = 1, 2, ..., M$, and $o \neq m$; and the $m-$th element of c_m is set to 1, i.e., $c_{mm} = 1, m = 1, 2, ..., M$. The rest of the elements in c_m are 0, that is, $c_{mo} = 0, o = 1, 2, ..., M$, and $o \neq m$. Except for the decision function $f(x)$, this method also needs to define a membership function $w(x)$, $w(x_i) \in \mathbb{R}^M$ for any instance x_i, and $w_m(x_i)$ is the membership of x_i belonging to the $m-$th class. Finally, through the improved classification method, each instance is constrained to share the same membership vector with its corresponding LWM according to the local learning principle (Bottou and Vapnik, 1992; Atkeson et al., 1997). The optimization problem of PCP-ER is formulated as

$$\begin{aligned} \min_{f, w_m(x_i)} & \sum_{m=1}^{M}\sum_{i=1}^{n} w_m(x_i)^b \|f(x_i) - c_m\|^2 \\ & + \lambda_s \sum_{m=1}^{M}\sum_{i=1}^{n} w_m(x_i)^b \|f(\hat{x}_i) - c_m\|^2 + \lambda\|f\|_{\mathcal{H}}^2 \\ & - C\sum_{m=1}^{M}\sum_{i=1}^{n}(-w_m(x_i)^b \ln w_m(x_i)^b + w_m(x_i)^b) \\ s.t. \quad & 0 \leq w_m(x_i) \leq 1, \quad m = 1, \ldots, M \end{aligned} \tag{2}$$

where λ_s, λ, and C are regularization parameters corresponding to each term in the objective function, and the parameter b is an exponent on label membership. b is used to control the uncertainty of instances belonging to multiple classes. Specifically, when $b = 1$, the value of each label membership $w_m(x_i)$ is taken from {0, 1}, which will cause PCP-ER to degenerate to the original clustering classification. That is, each instance belongs to only one class. When $b = \infty$, the label membership of all instances on all classes will be equal. In order to avoid the occurrence of trivial solutions, given $b = 2$ in the subsequent derivation process of this paper, the detailed proof process of the value of b has been given in Krishnapuram and Keller (1993). The first term of the objective function in Eq. (2) describes the minimization of the loss by a squared loss function; the second term describes the consistency between the predictions of each instance and its corresponding LWM by adjusting the parameter λ_s; the third term is a regularization term, which is used to prevent the model from over-fitting, and the complexity of the model is controlled by adjusting the parameter λ. The last term describes how to adjust the influence of noise on the model through the fuzzy entropy (Kosko, 1986; Krishnapuram and Keller, 1996), and $\sum_{m=1}^{M}\sum_{i=1}^{n}(-w_m(x_i)^b \ln w_m(x_i)^b + w_m(x_i)^b)$ calculates the fuzzy entropy. The larger the fuzzy entropy, the greater the amount of discriminative information of the sample. The model has better generalization ability. $-C\sum_{m=1}^{M}\sum_{i=1}^{n}(-w_m(x_i)^b \ln w_m(x_i)^b + w_m(x_i)^b)$ is a monotonically decreasing function about $w_m(x_i)$, and it needs to adjust the balance parameter C and force $w_m(x_i)$ to be as large as possible for avoiding trivial solutions. In addition, it can also make noise data have smaller different labels membership. Therefore, fuzzy entropy controls noise and outlier, making the method more robust, and its Robustness has been analyzed and proved in detail in (Krishnapuram and Keller, 1996).

For labeled instances, the label membership function is defined as

$$w_m(x_i) = \begin{cases} 1, \text{ if } x_i \in X_m, & \begin{matrix} i = 1, 2, ...l, \\ m = 1, 2, ...M, \end{matrix} \\ 0, \text{ else} \end{cases} \tag{3}$$

where X_m is a subset with instances belonging to the $m-$th class. The Eq. (2) can be rewritten as

$$\begin{aligned} \min_{f, w_m(x_j)} & \sum_{i=1}^{l}\|f(x_i) - y_i\|^2 + \lambda_s\sum_{i=1}^{l}\|f(\hat{x}_i) - y_i\|^2 \\ & + \sum_{m=1}^{M}\sum_{j=l+1}^{n} w_m(x_j)^2 \|f(x_j) - c_m\|^2 \\ & + \lambda_s\sum_{m=1}^{M}\sum_{j=l+1}^{n} w_k(x_j)^2 \|f(\hat{x}_j) - c_m\|^2 + \lambda\|f\|_{\mathcal{H}}^2 \\ & + C\sum_{m=1}^{M}\sum_{j=l+1}^{n}(w_m(x_j)^2 \ln w_m(x_j)^2 - w_m(x_j)^2) \\ s.t. \quad & 0 \leq w_m(x_j) \leq 1, \quad m = 1, \ldots, M, \; j = l+1, \ldots, n \end{aligned} \tag{4}$$

According to PCP-ER, each instance has a membership vector about all classes and each instance, and its corresponding LWM share the same membership vector.

It should be noted that, in Eq. (2), we adopt a square loss function. However, other classification loss functions can also be used to develop different semi-supervised classification methods based on the possibility clustering assumption. Compared with the Eq. (3) in SSCCM (Wang et al., 2012), the Eq. (2) relaxes the constraint that the sum of the label membership on all classes is 1, and it employs the fuzzy entropy to overcome the influence of noise and outlier for obtaining the more robust model.

OPTIMIZATION

The optimization problem of PCP-ER is a non-convex problem with regard to $f(x)$. We adopt an alternating iterative strategy to achieve the optimization of the decision function $f(x)$ and the label membership function $w(x)$ in this paper, and each iteration has a closed-form solution.

Fixed $w(x)$ firstly to optimize $f(x)$. Since the sixth term in Eq. (4) has no calculation for $f(x)$, we can get

$$\begin{aligned}\min_f & \sum_{i=1}^{l}\|f(x_i)-y_i\|^2+\lambda_s\sum_{i=1}^{l}\|f(\widehat{x_i})-y_i\|^2 \\ & +\sum_{m=1}^{M}\sum_{j=l+1}^{n}w_m(x_j)^2\|f(x_j)-c_m\|^2 \\ & +\lambda_s\sum_{m=1}^{M}\sum_{j=l+1}^{n}w_m(x_j)^2\|f(\widehat{x_j})-c_m\|^2+\lambda\|f\|_H^2 \\ s.t.\ & 0\le w_m(x_j)^2\le 1,\quad m=1,...,M,\quad j=l+1,...,n\end{aligned} \tag{5}$$

According to the Representer Theorem, the minimization problem of Eq. (5) exists in the Reproducing kernel Hilbert space (RKHS), and its solution form can be written as $f(x)=\sum_{i=1}^{n}a_iK(x_i,x)$ (Belkin et al., 2006). The minimization problem is simplified to optimize the finite-dimensional space of coefficient α_i. In this paper, Mercer kernel function $K=(K_{ij})\in\mathbb{R}^{(l+u)\times(l+u)}$, $K_{ij}=K(x_i,x_j)$, $K_i\in\mathbb{R}^{(l+u)\times l}$, $K_u\in\mathbb{R}^{(l+u)\times u}$, $K=[K_l\ K_u]=\begin{bmatrix}K_{ll}&K_{lu}\\K_{lu}^T&K_{uu}\end{bmatrix}$, $\overline{K}=[\overline{K}_l\ \overline{K}_u]=\begin{bmatrix}\overline{K}_{ll}&\overline{K}_{lu}\\\overline{K}_{lu}^T&\overline{K}_{uu}\end{bmatrix}$.

Theorem 1. The best solution of the original optimization problem of the Eq. (5) is $f(x)=\sum_{i=1}^{n}a_iK(x_i,x)$, where

$$\begin{aligned}\alpha=&(Y_lK_l^T+\hat{V}LJ^TK_u^T+\lambda_sY_l\bar{K}_l^T+\lambda_s\hat{V}LJ^T\bar{K}_u^T)\\&(K_lK_l^T+\hat{V}K_uJJ^TK_u^T+\lambda_s\bar{K}_l\bar{K}_l^T+\lambda_s\hat{V}\bar{K}_uJJ^T\bar{K}_u^T+\lambda K)^{-1}\end{aligned}$$

$\alpha_i\in\mathbb{R}^{M\times1}$, $\alpha=[\alpha_1,\alpha_2,...,\alpha_n]\in\mathbb{R}^{M\times n}$ is Lagrange multiplier matrix. $Y_l=(y_1,y_2,...,y_l)\in\mathbb{R}^{M\times l}$, $y_i\in\mathbb{R}^{M\times1}$, $i=1,2,...,l$, $J=[\underbrace{I_u,...,I_u}_{M}]\in\mathbb{R}^{u\times(M\times u)}$, $I_u\in\mathbb{R}^{u\times u}$ is an identity matrix, $L=[L_1,...,L_M]\in\mathbb{R}^{M\times(M\times u)}$, $L_m\in\mathbb{R}^{M\times u}$ is a matrix with all-one vector in the $m-$th row, the other being all-zero vectors. Let $V=[v(x_1)\dots v(x_u)]\in\mathbb{R}^{M\times u}$, $v(x_i)\in\mathbb{R}^{M\times1}$ refers to the membership values of each unlabeled instance on the M classes. $\widehat{V}$ is a diagonal matrix with each element on the diagonal correspond to the squared values of the elements in the corresponding row in the matrix V.

Proof. Like the Eq. (1), each LWM of x_i in the kernel space can be rewritten as. Then $K_{lu}=\langle\phi(X_l),\phi(X_u)\rangle_{\mathcal{H}}$, $K_{ll}=\langle\phi(X_l),\phi(X_l)\rangle_{\mathcal{H}}$, and $K_{uu}=\langle\phi(X_u),\phi(X_u)\rangle_{\mathcal{H}}$, where $\overline{K}_{ll}=\left\langle\phi(X_l),\widehat{\phi(X_l)}\right\rangle_{\mathcal{H}}$, $\overline{K}_{lu}=\left\langle\phi(X_l),\widehat{\phi(X_u)}\right\rangle_{\mathcal{H}}$, $\overline{K}_{ul}=\left\langle\phi(X_u),\widehat{\phi(X_l)}\right\rangle_{\mathcal{H}}$, and $\overline{K}_{uu}=\left\langle\phi(X_u),\widehat{\phi(X_u)}\right\rangle_{\mathcal{H}}$, each element $\overline{K}_{ij}$ can be formulated as

$$\begin{aligned}\bar{K}_{ij}&=\langle\phi(x_i),\widehat{\phi(x_j)}\rangle_{\mathcal{H}}=\left\langle\phi(x_i),\frac{\sum_{x_s\in Ne(x_j)}W_{sj}\phi(x_s)}{\sum_{x_s\in Ne(x_j)}W_{sj}}\right\rangle_{\mathcal{H}}\\&=\frac{\sum_{x_s\in Ne(x_j)}W_{sj}\langle\phi(x_i),\phi(x_s)\rangle_{\mathcal{H}}}{\sum_{x_s\in Ne(x_j)}W_{sj}}\\&=\frac{\sum_{x_s\in Ne(x_j)}W_{sj}K_{is}}{\sum_{x_s\in Ne(x_j)}W_{sj}}\end{aligned} \tag{6}$$

By the F-norm with $\|X\|_F^2=tr(X^TX)$, all matrices and $f(x)=\sum_{i=1}^{n}a_iK(x_i,x)$ can be substituted into Eq. (5), we have

$$\begin{aligned}\min_\alpha F_1=&tr\left((\alpha K_l-Y_l)(\alpha K_l-Y_l)^T\right)\\&+\lambda_s tr\left((\alpha\bar{K}_l-Y_l)(\alpha\bar{K}_l-Y_l)^T\right)\\&+tr\left((\alpha K_u\text{J-L})^T\hat{V}(\alpha K_u\text{J-L})\right)\\&+\lambda_s tr\left((\alpha\bar{K}_u\text{J-L})^T\hat{V}(\alpha\bar{K}_u\text{J-L})\right)+\lambda tr\left(\alpha K\alpha^T\right)\end{aligned} \tag{7}$$

Set the derivative of F_1 w.r.t. α to zero, we have

$$\begin{aligned}\partial F_1/\partial\alpha=&(\alpha K_l-Y_l)K_l^T+\hat{V}(\alpha K_uJ-L)J^TK_u^T\\&+\lambda_s(\alpha\bar{K}_l-Y)\bar{K}_l^T+\lambda_s\hat{V}(\alpha\bar{K}_uJ-L)\\&J^T\bar{K}_u^T+\lambda\alpha K=0.\end{aligned} \tag{8}$$

According to the Eq. (8), we can get the solution of α, Theorem 1 is proved.

By Fixing $f(x)$ to optimize $w(x)$ from Eq. (4), we can obtain

$$\begin{aligned}\min_{w_m(x_j)}F_2=&\sum_{m=1}^{M}\sum_{j=l+1}^{n}w_m(x_j)^2\|f(x_j)-c_m\|^2\\&+\lambda_s\sum_{m=1}^{M}\sum_{j=l+1}^{n}w_m(x_j)^2\|f(\hat{x}_j)-c_m\|^2\\&+C\sum_{m=1}^{M}\sum_{j=l+1}^{n}(w_m(x_j)^2\ln w_m(x_j)^2-w_m(x_j)^2)\\s.t.\quad&0\le w_m(x_j)\le1,\quad m=1,\dots,M,\\&j=l+1,\dots,n\end{aligned} \tag{9}$$

Theorem 2. The optimal solution of the original optimization problem of the Eq. (4) is

$$w_m(x)=\exp\left(\frac{-(\|f(x)-c_m\|^2+\|f(\hat{x})-c_m\|^2)}{2C}\right). \tag{10}$$

Proof. Set the derivative of F_2 w.r.t. $w_m(x_j)$ to zero, we have

$$\begin{aligned}\partial F_2/\partial w_m(x_j)=&2w_m(x_j)\|f(x_j)-c_m\|^2\\&+2\lambda_s w_m(x_j)\|f(\hat{x}_j)-c_m\|^2\\&+C[2(w_m(x_j)\log w_m(x_j)^2]=0\end{aligned} \tag{11}$$

The solution of $w_m(x_j)$ is

$$w_m(x_j)=\exp\left(\frac{-(\|f(x_j)-c_m\|^2+\|f(\hat{x}_j)-c_m\|^2)}{2C}\right). \tag{12}$$

Therefore, the label membership vector of any instance x can be derived from Eq. (10), and Theorem 2 is proved.

ALGORITHM DESCRIPTION

The optimization of PCP-ER adopts an alternating iterative strategy. PCP-ER belongs to the category of semi-supervised large boundary methods that directly seek large boundary separators. In fact, iterative learning processes are often used in various semi-supervised learning methods. The initial value of membership of an unlabeled instance can be obtained through several strategies, such as randomization strategies, fuzzy clustering techniques (such as FCM), or all zeros simply being set. Actually PCP-ER is start with labeled data to initialize the decision function $f(x)$ in this paper. When $|F(\alpha_m, w_m(x)) - F(\alpha_{m-1}, w_{m-1}(x))| < \varepsilon F(\alpha_{m-1}, w_{m-1}(x))$, the iteration terminates, $F(\alpha_m, w_m(x))$ is the value of the objective function at the $m-$th iteration, and ε is a iterative termination parameter.

ALGORITHM DESCRIPTION OF PCP-ER

Input: the labeled data X_l with the labels Y_l, the unlabeled data X_u, the regularization parameters λ, λ_s, C, the iterative termination parameter ε, and the iterative maximum times T.

Output: the decision function $f(x)$, the label membership function $w(x)$.

Procedure:

1. Initialize the label memberships of unlabeled data;
2. Obtain the initial α by Eq. (6);
3. Obtain the initial $w(x)$ by Eq. (11);
4. Calculate the $F(\alpha_0, w_0(x))$ of objective function

for $m = 1$ **to** T **do**

{

5.1 Update α by Eq. (6);

5.2 Update $w(x)$ by Eq. (11);

5.3 Update the objective function value $F(\alpha_m, w_m(x))$;

5.4 **if** $|F(\alpha_m, w_m(x)) - F(\alpha_{m-1}, w_{m-1}(x))| < \varepsilon F(\alpha_{m-1}, w_{m-1}(x))$

break; **return** $f(x)$ and $w(x)$;

endif

endfor

}

DISCUSSION

PCP-ER Reliability

It takes the decision function and label membership function to identify each other's predicted classification results in order to further enhance the reliability of PCP-ER. This leads to Theorem 3.

Theorem 3. The decision function and label membership function are adopted in PCP-ER to obtain predictions, and their predictions are usually consistent (actually consistent or indirectly consistent). If the two predictions are not consistent, the predict instance may be located near the decision boundary and these predictions may be unreliable.

Proof. Each instance can be predicted by the decision function $y^* = \arg\max_{m=1,\ldots,M} f_m(x)$ from Theorem 1 or the label membership function $y^* = \arg\max_{m=1,\ldots,M} w_m(x)$ from Theorem 2. In the case of $\forall j = 1, \ldots, M; j \neq m$, if $f(x)$ is used to predict x and $f_m(x) > f_j(x)$, then $x \in X_m$. If $w(x)$ is taken to predict x, $w_m(x) > w_j(x)$, the result of $x \in X_m$ can also be obtained. If λ_s is fixed, $f_m(x) + \lambda_s f_m(x) > f_j(x) + \lambda_s f_j(x)$ can also obtain the above consistent prediction result. When $\lambda_s = 0$, the predictions of $f(x)$ and $w(x)$ are consistently. When $\lambda_s \neq 0$, x, and $\widehat{x}$ share the same label from $f(x)$, that is, $\arg\max_{m=1,\ldots,M} f_m(x) = \arg\max_{m=1,\ldots,M} f_m(\widehat{x})$, the prediction results of $f(x)$ and $w(x)$ are also consistent. If $f_j(\hat{x}) - f_m(\hat{x}) < \left(f_m(x) - f_j(x)\right)/\lambda_s \forall j = 1, \ldots, M, j \neq m$, the prediction results of $f(x)$ and $w(x)$ are also consistent. If x is located near the decision boundary, the prediction is obviously different between x and $\hat{x}$, and it is possible that $\hat{x}$ and x are located in different class, then this prediction of x is unreliable.

Finally, three instances can be summarized:

(1) Intrinsic consistent instance, where instances x and $\hat{x}$ get the same label by $f(x)$, then the prediction results of $f(x)$ and $w(x)$ on x are consistent;
(2) Fake-consistent instance, where x is not an intrinsic consistent instances, but $f_j(\hat{x}) - f_m(\hat{x}) < \left(f_m(x) - f_j(x)\right)/\lambda_s$ and the prediction results of $f(x)$ and $w(x)$ on x are still consistent;
(3) Inconsistent instance, where the prediction results of $f(x)$ and $w(x)$ on x are not consistently.

Thus this theorem is proved.

Actually, only one function is needed to predict new instances, and if the memberships of some instances are expected to be obtained, the label membership function is preferred. If these two functions are taken to predict instances at the same time, their prediction inconsistency is used to detect those boundary instances that are difficult to classify, and we do some special processing on them, such as manual labeling, to improve classification reliability. The prediction of these two functions can verify each other, and the reliability of semi-supervised classification can be enhanced by checking their consistency.

PCP-ER Convergence

In order to prove the convergence of Algorithm 1, we have Theorem 4.

Theorem 4. The sequence $\{F(\alpha_m, w_m)\}$ obtained from the above algorithm is convergent.

Proof. Since the objective function $F(\alpha, w)$ is a biconvex function on (α, w) (Gorski and Pfeuffer, 2007). Fixed $w(x)$ and the objective function is a convex function on α, so the optimal α^* can be calculated by minimizing $F(\alpha, w_m)$ in Eq. (6) or optimizing Eq. (5) equivalently. Given $\alpha_{m+1} = \alpha^*$, we know that $F(\alpha_{m+1}, w_m) = F(\alpha^*, w_m) \leq F(\alpha_m, w_m)$. At this time, Fixed α_{m+1} and the objective function is

a convex function on w. Therefore, the optimal w can be obtained by minimizing $F(\alpha_{m+1}, w_m)$ in Eq. (10) or optimizing Eq. (9) equivalently. Given $w_{m+1} = w^*$, we know that $F(\alpha_{m+1}, w_{m+1}) = F(\alpha_{m+1}, w^*) \leq F(\alpha_{m+1}, w_m)$, it can be inferred that$F(\alpha_{m+1}, w_{m+1}) \leq F(\alpha_{m+1}, w_m) \leq F(\alpha_m, w_m)$, $\forall m \in N$, $\{F(\alpha_m, w_m)\}$ is monotonically decreasing. Since the objective function is non-negative, it has a lower bound. Thus the theorem is proved.

Generalization Error Bound of PCP-ER

In statistical learning theory, the VC dimension (Vapnik Chervonenkis dimension; Vapnik, 1995) provided a generalization error bound method that can analyze machine learning (Bishop, 2006). Therefore, this paper selects the VC dimension method to analyze the generalization error bound of PCP-ER.

Theorem 5. (PCP-ER generalization error bound) Let H be the RKHS. The generalization error bound of the learning function $f^{\Phi} \in H$ that satisfies the following Eq. (13) at the probability $1 - \delta$ $(0 < \delta < 1)$:

$$\begin{aligned} R\left(f^{\Phi}\right) &\leq \varepsilon + \frac{1}{l}\sum_{i=1}^{l} L(f^{\phi}(x_i), y_i) + \\ &\quad \frac{1}{n-l}\sum_{m=1}^{M}\sum_{i=l+1}^{n} w_m^2(x_i)\, L(f^{\phi}(x_i), c_m), \\ &\leq \varepsilon + \frac{1}{n}\sum_{i=1}^{n} L(f^{\phi}(x_i), y_i) \end{aligned} \tag{13}$$

where $R\left(f^{\Phi}\right)$ is the expected error, it is the PCP-ER generalization error; the right side of the inequality is the upper bound of the generalization error, $\varepsilon = \sqrt{\frac{1}{2n}\left(\ln d + \ln \frac{1}{\delta}\right)}$ is a constant, d is the number of functions in the hypothesis space, n is the number of samples, and δ is the probability of a function occurrence in the hypothesis space. Since $\frac{1}{n}\sum_{i=1}^{n} L(f(x_i), y_i)$ is the empirical error of the traditional semi-supervised classification method. According to the design idea of PCP-ER, $\frac{1}{l}\sum_{i=1}^{l} L(f(x_i), y_i)$ is the empirical risk of labeled data and $\frac{1}{n-l}\sum_{m=1}^{M}\sum_{i=l+1}^{n} w_m^2(x_i)\, L(f(x_i), c_m)$ the empirical risk of unlabeled data in (14), and $0 \leq w_m(x_i) \leq 1$, $m = 1, \ldots, M$; $i = l+1, \ldots, n$. It follows that the second inequality in (13) is true. Compared with the empirical error of traditional semi-supervised classification methods, PCP-ER has a smaller generalization error bound and a better generalization.

By analysis of Theorem 5, the generalization error of this method can be adjusted and controlled by the membership function $w_m(x)$, which makes it possible to obtain better generalization.

EXPERIMENT

This section will compare the PCP-ER method with the latest semi-supervised recognition method, the hard PCP-ER method, the PCP-ER method joint features by multiple kernel functions (called as MKPCP-ER for short), the PCP-ER method with the features from deep learning (called as DLPCP-ER for short), respectively. The emotion recognition results of all these methods are compared on the real EEG-based datasets [i.e., DEAP (Koelstra et al., 2012), SEED (Zheng and Lu, 2015), and SEED-IV (Zheng et al., 2019)]. It aims to study the following three issues:

(1) How does PCP-ER compare to the latest semi-supervised classification methods?
(2) How does PCP-ER compare to MKPCP-ER, DLPCP-ER and hard PCP-ER?
(3) How does the regularization parameter$\lambda_s$affect the intrinsic consistence of PCP-ER?

Datasets Description

There are a few existing EEG datasets that can be used for affective states investigation. In this paper, we use three publicly available datasets: DEAP (Koelstra et al., 2012), SEED (Zheng and Lu, 2015), and SEED-IV (Zheng et al., 2019).

The DEAP dataset contains 32 experimental subjects. While recording physiological signals, each subject needs to watch 40 1-min music videos as emotional stimuli. The resulting dataset includes 32-channel EEG signals, 4-channel electroencephalogram, 4-channel electromyogram, respiration, plethysmograph, Galvanic Skin Response, and body temperature. Each subject recorded 40 EEG trials, each trial corresponding to an emotion caused by a music video. After watching each video, subjects need to immediately evaluate their truly-felt emotion from five dimensions: valence (related to pleasantness level), arousal (related to excitation level), dominance (related to control power), liking (Related to preference), and familiarity (related to stimulating knowledge). The rating ranges from 1 (weakest) to 9 (strongest) except for familiarity, which is rated from 1 to 5. The EEG signals were recorded by a Biosemi Active Two device at a sampling rate of 512 Hz and down-sampled to 128 Hz.

The original SEED dataset contains 15 experimental subjects. The movie clips are intended to elicit three emotions—positive, neutral, and negative emotions—and five movie clips are assigned to each emotion. All subject need to experience three EEG recording sessions, with two consecutive recording experiments separated by 2 weeks. In each experiment, each subject watch 15 movie clips, each of which was about 4 min long, to induce the desired emotions. The same 15 movie clips were used in all three experiments. The result dataset contains 15 EEG trials for each subject in each experiment and 5 trials for each emotion. EEG signals (EEG signals) are recorded by 62-channel[1] ESI NeuroScan equipment, with a sampling rate of 1,000 Hz and down-sampling to 200 Hz.

For the SEED-IV dataset, a total of 168 movie clips in the material pool containing four emotions (happy, sad, fear, and neutral), 44 participants (22 women, college students) were asked to evaluate their emotional state when watching the

[1]The 62 EEG channels include AF3, AF4, C1, C2, C3, C4, C5, C6, CB1, CB2, CP1, CP2, CP3, CP4, CP5, CP6, CPZ, CZ, F1, F2, F3, F4, F5, F6, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCZ, FP1, FP2, FPZ, FT7, FT8, FZ, O1, O2, OZ, P1, P2, P3, P4, P5, P6, P7, P8, PO3, PO4, PO5, PO6, PO7, PO8, POZ, PZ, T7, T8, TP7, and TP8.

movie clips [Scoring in the two dimensions of valence and arousal; -5 ∼ 5)]. The valence scale ranges from sad to happy. Arousal is measured from calm to excited. Finally, 72 movie clips with the highest recognition are carefully selected from the material pool for the four emotion-evoking experiments. Each movie clip is about 2 min long. The experiment contains 15 experimental subjects. In order to investigate the stability of the model over time, each experimental subject needs to make a total of three experimental records in different periods, to avoid repetition, each movie clip is used for only one trial by the same subject during the three experimental periods. Each experimental for one subject contains 24 trials (each 6 trials correspond to one emotion). The resulting dataset consists of 45 experiments for all subjects. A 62-channel ESI NeuroScan System was used to record the experimenter's EEG signal. The sampling rate was 1,000 Hz, and the sample was down-sampled to 200 Hz. SMI eye-tracking glasses records the eyes movement.

Baselines

This experimental part compares the PCP-ER method in this paper with LapSVM (Belkin et al., 2006), LapRLS (Belkin et al., 2006), TSVM (Joachims, 1999)[2], meanS3VM (Li et al., 2009)[3], and SSCCM (Wang et al., 2012)—five newest semi-supervised classification methods.

LapSVM is the Laplacian Support Vector Machine. This method takes the manifold hypothesis for semi-supervised classification. The loss function is the hinge loss function. According to the Laplacian graph, it searches a maximum-face decision function the entire data distribution.

LapRLS is the Laplacian regularized least squares. The method also uses the manifold hypothesis for semi-supervised classification, but the loss function is the least square loss function.

TSVM is the Transduced Support Vector Machine. This method uses the clustering assumption, in order to find an interface on labeled and unlabeled data, so as to guide the classification boundary through low-density regions.

MeanS3VM is a type of semi-supervised SVM based on the mean value of unlabeled data. The clustering assumption is also adopted, which actually contains two implementation methods (Krishnapuram and Keller, 1996), namely the meanS3VM-iter method based on alternating optimization and the meanS3VM-mkl method based on multiple kernel learning.

Semi-supervised classification method based on modified cluster assumption is a new semi-supervised classification method based on modified clustering assumption. The clustering assumption is also used, its purpose is to find a membership function and a decision function on labeled and unlabeled data, so that similar instances should share similar label membership, and one instance can belong to multiple Classes.

Besides, we then also compare the PCP-ER method with the hard PCP-ER, which only uses the hard clustering assumption. It is supposed in hard PCP-ER that each instance clearly belongs to only one class, which can be formulated in Eq. (14), where each $c_i \in \mathbb{R}^M$ is an one-hot vector.

$$\begin{aligned}\min_{f,y_j} \ & \sum_{i=1}^{l} \|f(x_i) - y_i\|^2 + \lambda_s \sum_{i=1}^{l} \|f(\hat{x}_i) - y_i\|^2 \\ & + \sum_{j=l+1}^{n} \|f(x_j) - y_j\|^2 + \lambda_s \sum_{j=l+1}^{n} \|f(\hat{x}_j) - y_j\|^2 . \\ & + \lambda \|f\|_{\mathcal{H}}^2 \\ s.t. \quad & y_j \in \{c_1, ..., c_M\}, \quad j = l+1, \ldots, n\end{aligned} \tag{14}$$

The objective function in Eq. (14) can also be readily solved by the same strategy as adopted in our PCP-ER.

Experimental Setting

In this section, we will give the experimental settings of the three datasets on the compared methods.

The DEAP dataset is a three binary classification (including valence, arousal, and dominance). One experimental subject contributed 40 samples. Then, the training set includes $40 \times 32 = 1{,}240$ samples from 32 experimental subject, and the test set contains 40 test samples from test subjects. In the DEAP training set, there are three settings: the first one contains 10 labeled instances, the second one contains 50 labeled instances, and the third one contains 100 labeled instances. Furthermore, each setting is related to 12 subsets of labeled data and average performance results on unlabeled data. As with other machine learning algorithms, these state-of-the-art semi-supervised classification algorithms require certain hyper-parameters to be set. For some hyper-parameters, we set them to the default values suggested by their authors. **Table 1** gives details of the other hyper-parameters used in this experiment. Here, you need to use a linear kernel and an RBF kernel. When 10 labeled instances are provided, the width parameter in the RBF kernel is set to the average distance between the instances, and when there are 50 labeled instances, leave-one-subject-out cross-validation is to be taken over the labeled data. When there are 100 labeled instances, 10-fold cross-validation is used on the labeled data, and finally a better result is selected between these two kernels.

The SEED dataset is a three-category dataset, including negative, neutral, and positive. One subject contributes 45 samples. The training set includes $45 \times 14 = 630$ samples from 14 subjects, a total of 210 samples per session, and 45 samples from the test subject. The training data contains 10 labeled and rest unlabeled instances. The linear kernel and a leave-one-subject-out cross-validation method a taken during this process to evaluate emotion recognition accuracy. As with other machine learning algorithms, these state-of-the-arts algorithms that require certain hyper-parameters to be set. For some hyper-parameters, we set them to the default values suggested by their authors. **Table 2** gives details of the hyper-parameters used in this experiment. C is used here to describe the divergence of the dataset. Different C values can be obtained according to different datasets, and $\bar{x} = \frac{\sum_{j=1}^{n} x_j}{n}$.

[2]The codes of LapSVM, LapRLS and TSVM are available from https://www.csie.ntu.edu.tw/~cjlin/libsvm/

[3]The code is downloaded from http://www.lamda.nju.edu.cn/CH.Data.ashx

TABLE 1 | Details of hyper-parameters on SEED dataset.

Method	Hyper-parameters
LapSVM	$C_1 = 1, C_2 = 0.1$
LapRLS	See above
TSVM	See above
MeanS3VM-iter	See above
MeanS3VM-mkl	See above
SSCCM	$\lambda = 0.1,\ \lambda_s = 0.1,\ m = 5,\ \varepsilon = 10^{-3}, C = \frac{\sum_{j=1}^{n} \|x_j - \bar{x}\|}{n}$,
PCP-ER	See above
Hard PCP-ER	See above

TABLE 2 | Details of hyper-parameters on SEED-IV dataset.

Method	Hyper-parameters
LapSVM	$C_1 = 1, C_2 = 0.1$
LapRLS	See above
TSVM	See above
MeanS3VM-iter	See above
MeanS3VM-mkl	See above
SSCCM	$\lambda = 0.1,\ \lambda_s = 0.1,\ m = 5,\ \varepsilon = 10^{-3}, C = \frac{\sum_{j=1}^{n} \|x_j - \bar{x}\|}{n}$,
PCP-ER	See above
Hard PCP-ER	See above

TABLE 3 | Details of hyper-parameters on DEAP dataset.

Method	Hyper-parameters
LapSVM	$C_1 = 100, C_2 = 0.1$
LapRLS	See above
TSVM	See above
MeanS3VM-iter	See above
MeanS3VM-mkl	See above
SSCCM	$\lambda = 1, \lambda_s = 0.1, m = 5, \varepsilon = 10^{-3}$
Hard PCP-ER	See above
PCP-ER	See above

The SEED-IV dataset is a four-category dataset: happy, sad, neutral, and fear. One subject contributes 72 samples. The training set includes $72 \times 14 = 1{,}008$ samples from 14 subjects; there are three sessions, a total of 336 samples per session, and 72 samples from the test subject. The training data contains 10 labeled instances, and the others are unlabeled instances. This process uses a linear kernel and a leave-one-out cross-validation method to evaluate classification accuracy. As with other machine learning algorithms, these state-of-the-arts algorithms that require certain hyper-parameters to be set. For some hyper-parameters, we set them to the default values suggested by their authors. **Table 3** gives details of the hyper-parameters used in this experiment. C is used here to describe the divergence of the data set. Different C values can be obtained according to different datasets, and $\bar{x} = \frac{\sum_{j=1}^{n} x_j}{n}$.

TABLE 4 | Performance about PCP-ER and the latest methods on DEAP dataset.

Methods	DEAP		
	10 Labels	**50 Labels**	**100 Labels**
LapSVM	49.22	53.05	63.22
LapRLS	50.06	57.49	63.46
TSVM	44.70	47.66	52.49
MeanS3VM-iter	49.83	53.43	59.17
MeanS3VM-mkl	52.11	58.47	60.54
SSCCM	**56.54**	61.31	63.33
PCP-ER	55.7	**62.40**	**66.71**
Consis. rate	0.9827	0.9943	1.00

Experimental Results and Analysis

Specifically, in the following tables (i.e., **Tables 4**, **5**) of experimental results, the bold values in each column indicate the best accuracy in all these tables, and the bold values in last column from **Table 5** indicate the best average performance results achieved by the compared methods. The last Avg. column shows the average performance of each method on all data sets. In **Tables 4**, **5**, the consistency rate of PCP-ER on different settings of dataset is given in the last row. From the results of **Tables 4**, **5**, we can draw the following conclusions.

Results on DEAP

The experimental comparison among PCP-ER and the six latest methods on DEAP with different number of labeled samples. The experimental results are shown in **Table 4**. The consistency rate gradually approaches to 1 which increases as the labeled samples increase. Since the SSCCM is used to deal with normal data, it obtains the best performance with 10 labeled samples. The possible reason is that the samples are normal. However, the SSCCM method is slightly better than the PCP-ER method in this case, and the PCP-ER method has the best performance in most cases with the increase of the number of labeled samples. It shows that the clustering hypothesis with fuzzy entropy can overcome the influence of noise and outliers in semi-supervised classification.

Results on SEED and SEED-IV

We performed an experimental comparison among PCP-ER and the six latest methods on each session with 10 labels of SEED and SEED-IV datasets. The experimental results are shown in **Table 5**. The average performance of the proposed method is the best. In addition, the consistency rate is close to 1. Although the consistency rate on Session 2 of SEED-IV dataset isn't good, its value is as high as 0.988. It is worth noting that only method MeanS3VM-mkl performs slightly better than PCP-ER method on session 2 of SEED dataset, the possible reason is that the MeanS3VM-mkl employed multiple kernels and enriched features of dataset. Nevertheless, the performance of the PCP-ER method is also closely followed, and the proposed method has the best performance of all the other datasets. It shows that the clustering hypothesis with fuzzy entropy can overcome the influence of noise in semi-supervised emotion recognition on SEED and SEED-IV.

TABLE 5 | Performance comparison among PCP-ER and the latest methods on SEED and SEED-IV datasets.

Methods	Seed				Seed-IV			
	Session1	Session2	Session3	Avg.	Session1	Session2	Session3	Avg.
LapSVM	52.26	49.46	58.39	53.37	58.00	51.88	56.33	56.33
LapRLS	52.08	50.55	57.16	53.26	57.29	51.04	55.72	55.72
TSVM	49.83	47.29	53.44	50.19	55.73	45.20	50.60	50.6
MeanS3VM-iter	55.27	49.71	58.21	54.40	58.08	48.95	56.49	56.49
MeanS3VM-mkl	60.02	**51.20**	58.78	56.67	60.29	51.68	57.25	57.25
SSCCM	61.78	50.68	60.11	57.52	59.38	51.17	61.78	61.78
PCP-ER	**62.13**	50.04	**60.48**	**57.55**	**62.45**	**52.30**	**62.36**	**62.36**
Consis. rate	0.991	0.99	1.00	0.994	1.00	0.988	0.999	0.999

Multiple-Kernel Learning

We further evaluate the effectiveness of our method with different kernel functions (called as MKPCP-ER for short) to present instances from each dataset. Given the empirical kernel mapping set $\{\phi_k\}_{k=1}^{\mho}$, each mapping X_a into $\mho$ different kernel spaces, each X corresponds to each dataset [i.e., DEAP (Koelstra et al., 2012), SEED (Zheng and Lu, 2015), and SEED-IV (Zheng et al., 2019)], and we can integrate them orthogonally to the final space by concatenation, i.e., $\widetilde{\phi}(x_i) = [\phi_1(x_i)^T, \phi_2(x_i)^T, ..., \phi_{\mho}(x_i)^T]^T \in \mathbb{R}^{\mho n_a}$, for $x_i \in X_a$, $n_a (a = 1, 2, 3)$ training samples in each dataset. The final kernel matrix in this new space is defined as $K_{new} = [\widetilde{K}_1; \widetilde{K}_2; ...; \widetilde{K}_{\mho}]$, where $\widetilde{K}_i$ is the kernel matrix in the $i-$thfeature space about $i-$th dataset. Therefore, besides the above mentioned Gaussian kernel, we additionally employ another three types of kernels in MKPCP-ER: Laplacian kernel $K_{ij} = \exp\left(-\sqrt{\sigma}\|x_i - x_j\|\right)$, inverse square distance kernel $K_{ij} = 1/\left(1 + \sigma\|x_i - x_j\|^2\right)$, and inverse distance kernel $K_{ij} = 1/\left(1 + \sqrt{\sigma}\|x_i - x_j\|\right)$.

It can be clearly seen from **Figure 2** that MKPCP-ER is obviously better than PCP-ER in terms of mean accuracies in all cases, which justifies that the multi-kernel trick can improve the quality of semi-supervised emotion recognition on each dataset.

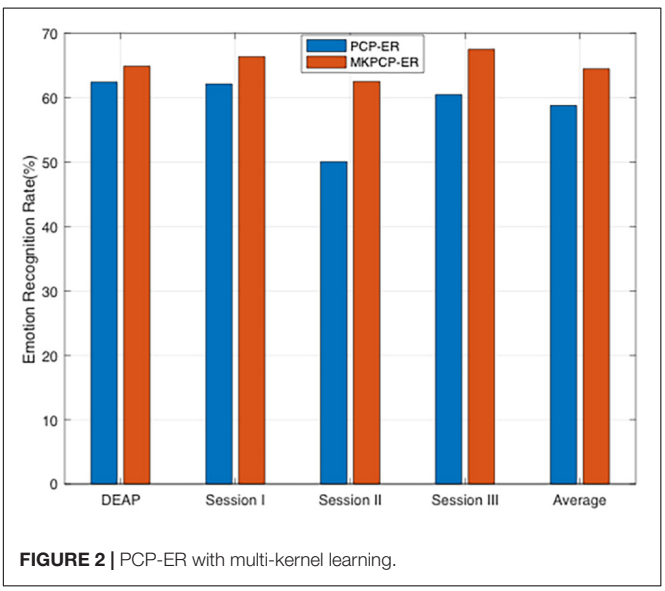


FIGURE 2 | PCP-ER with multi-kernel learning.

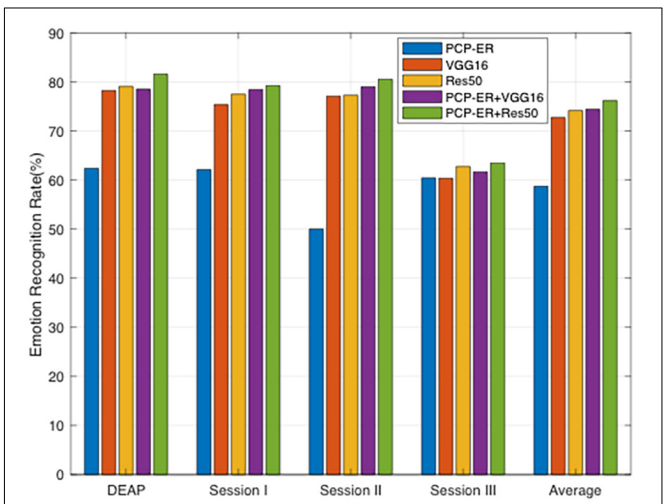


FIGURE 3 | Emotion recognition accuracies (%) of different methods using deeply extracted features.

Deep Features Learning

In the past decades, deep learning attracts more and more attention due to its powerful representation ability and dramatic improvement over the traditional shallow methods. The EEG emotion recognition method based on deep learning has also been widely used and has achieved better recognition effect than traditional methods. For example, in Zheng and Lu (2015) proposed a deep belief network for EEG emotion classification; in Song et al. (2018) and Zhong et al. (2020), authors used graphics to model multi-channel EEG features and then classified EEG emotion on this basis; Li et al. (2018) proposed a new neural network model, which uses time information for EEG emotion

TABLE 6 | Performance comparison between hard PCP-ER and PCP-ER on DEAP dataset.

Methods	DEAP			
	10 Labels	50 Labels	100 Labels	Avg.
hard PCP-ER	53.27	58.44	62.69	58.1
PCP-ER	**55.7**	**62.40**	**66.71**	**61.1**

TABLE 7 | Performance comparison between hard PCP-ER and PCP-ER on SEED and SEED-IV datasets.

Methods	SEED				SEED-IV			
	Session1	Session2	Session3	Avg.	Session1	Session2	Session3	Avg.
hard PCP-ER	60.57	47.32	56.78	54.89	60.38	49.08	61.66	57.04
PCP-ER	**62.13**	**50.04**	**60.48**	**57.55**	**62.45**	**52.30**	**62.36**	**59.03**

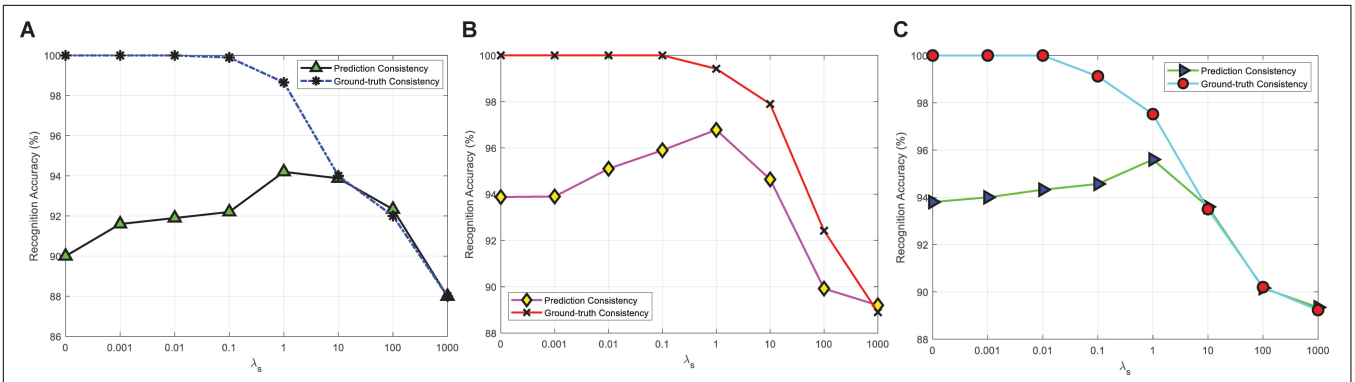


FIGURE 4 | Prediction consistency and ground-truth consistency between $f(x)$ and $w(x)$ with different λ_S values on **(A)** DEAP; **(B)** SEED; and **(C)** SEED-IV.

recognition task. We therefore additionally compare our PCP-ER method with the recently proposed deep transfer learning models VGG16 (Simonyan and Zisserman, 2014) and ResNet50 (He et al., 2016; it is the same as Res50 in the following) for emotion recognition using deeply extracted features. In our PCP-ER, we can tackle the problem of deep emotion recognition with two steps: firstly, a higher-level feature extraction is learnt in an unsupervised fashion from all available datasets using the popular deep architectures (e.g., VGG16 or Res50); secondly, our PCP-ER is trained on the transformed data of all datasets and then used to classify test dataset. For fair comparison, however, we follow the experimental setup in Zhou et al. (2018) and Zhu et al. (2017). Specifically, we first fine-tune pre-trained deep models (e.g., VGG16, and Res50) by using the labeled samples in the dataset, and then use these fine-tuned CNN models to extract the features from EEG in dataset. Finally, we perform emotion recognition using PCP-ER on these deeply extracted features. In the context of our experiments, we denote our methods with different deep models as PCP-ER+VGG16, PCP-ER+Res50, respectively. As for VGG16 and Res50, we use their released source codes and fine-tune the pre-trained deep models, respectively.

All experimental results are reported in **Figure 3**. As can be seen from this plot, the deep learning methods are originally proposed to learn each dataset features, while our proposed method aims to improve the anti-interference ability, namely, their methods focus on feature learning, while our work focuses on emotion recognition. So our proposed method can be used to further improve the recognition accuracies by employing the features extracted by deep models, i.e., VGG16 and Res50, rather than the original EEG patterns. This indicates that the recognition-level constraint can preserve all discriminative structures of labeled samples for the guidance of unlabeled samples recognition, which demonstrates the effectiveness of PCP-ER framework. From the plot bars of **Figure 3**, it can be observed that PCP-ER+VGG16 consistently outperforms VGG16, while PCP-ER+Res50 is consistently better than Res50, which demonstrates that our PCP-ER method is complementary to the two deep learning methods VGG16 and Res50 by exploiting more features from dataset.

Comparison Between PCP-ER and Hard PCP-ER

Tables 6, **7** show the performance results of PCP-ER and hard PCP-ER on DEAP, SEED, and SEED-IV datasets, respectively. Specifically, in the following tables (i.e., **Tables 6**, **7**) of experimental results, the bold values in each column indicate the best accuracy, and the bold values in last column of different dataset indicate the best average performance results achieved by the compared hard PCP-ER and PCP-ER method. From the overall observation of the results in **Tables 6**, **7**, when there are more training samples or more labeled samples, the PCP-ER and hard PCP-ER methods are better. Moreover, the performance and average performance of PCP-ER are better than hard PCP-ER. It demonstrates that the PCP-ER method based on membership degree is effective and robust on EEG-based emotion recognition.

Consistency Analysis

Figure 4 shows the experimental results of PCP-ER actual consistency rates corresponding to different λ_s values {0, 0.001, 0.01, 0.1, 1, and 101001000} on DEAP, SEED, SEED-IV three datasets. In **Figures 4A,C**, when λ_s is small enough, the prediction consistency rate can reach 100%, and then the consistency rate gradually decreases with the increase of λ_s. Since the indirect consistency instance becomes the inconsistent instance, it finally becomes equal to the ground-truth consistency rate. In **Figure 4B**, when λ_s is in the range of 1 to 1000, the prediction consistency rate and the ground-truth consistency rate are not equal, although the trend of change is the same with

Figures 4A,C, and the possible reason is that there are the least training samples on SEED dataset. In addition, in **Figures 4A–C**, with the increase of λ_s to 1, the ground-truth consistency rate also increases. When λ_s continues to increase, the ground-truth consistency rate begins to decrease. It may be that when λ_s is far less than or greater than 1, the PCP-ER will focus more on the emotion recognition of samples or LWM samples and not on its prediction consistency.

CONCLUSION

The existing GSSL methods construct undirected weight grapy, which is sensitive or not enough robust to noise or outlier from EEG patterns. At the end, we proposed PCP-ER for EEG-based affective recognition. By adding the regularization term of fuzzy entropy, the amount of discrimination information of samples is increased, and a more generalized emotion recognizer is obtained through learning to overcome the negative effects of noise and outliers and improve the robustness of the method. Experimental results on the three real datasets DEAP, SEED, and SEED-IV show that the proposed method improves more the reliability and robustness of emotion recognition than competing algorithms. Moreover, both PCP-ER with multi-kernel and depth features extraction obtained better performance than PCP-ER. These tricks can improve the quality of the PCP-ER method on each dataset. However, in the process of optimization, how to obtain an effective combined multi-kernel function or kernel space and how to analyze and demonstrate the consistency of the proposed method in theory are issues worthy of further discussion.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported partly by two High level talents introduction research projects of Ningbo Polytechnic (under grant nos. RC201808 and RC201702), and partly by Foundation of Zhejiang Educational Committee (under grant no. Y201941140), and partly by Zhejiang basic public welfare research program (under grant no. LGG20F020013).

REFERENCES

Atkeson, C.-G., Moore, A.-W., and Schaal, S. (1997). Locally weighted learning. *Artif. Intell. Rev.* 11, 11–73. doi: 10.1007/978-94-017-2053-3_2

Belkin, M., Niyogi, P., and Sindhwani, V. (2006). Manifold regularization: a geometric framework for learning from examples. *J. Mach. Learn. Res.* 7, 2399–2434.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Bottou, L., and Vapnik, V. (1992). Local learning algorithms. *Neural Comput.* 4, 888–900. doi: 10.1162/neco.1992.4.6.888

Chapelle, O., Bernhard, S., and Alexander, Z. (2006). *Semi-Supervised Learning*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262033589.001.0001

Chu, W. S., De la Torre, F., and Cohn, J. F. (2017). Selective transfer machine for personalized facial action unit detection. *IEEE Trans. Patt. Analys. Mach. Intellig.* 39, 529–545. doi: 10.1109/TPAMI.2016.2547397

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194. doi: 10.1126/science.1076358

Gao, J., Wang, S.-T., and Deng, Z.-H. (2010). Global and local preserving based semi-supervised support vector machine. *Acta Electron. Sin.* 38, 1626–1633.

Gorski, J., and Pfeuffer, F. (2007). Biconvex sets and optimization with biconvex functions: a survey and extensions. *Math. Methods Operation Res.* 66, 373–407. doi: 10.1007/s00186-007-0161-1

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 770–778. doi: 10.1109/CVPR.2016.90

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/taffc.2014.2339834

Joachims, T. (1999). "Transductive inference for text classification using support vector machines," in *Proceedings of the 16th International Conference on Machine Learning*, Bled, 200–209.

Kim, M.-K., Kim, M., Oh, E., and Kim, S.-P. (2013). A review on the computational methods for emotional state estimation from the human EEG. 2013:573734. doi: 10.1155/2013/573734

Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/t-affc.2011.15

Kosko, B. (1986). Fuzzy entropy and conditioning. *Inform. Sci.* 40, 165–174. doi: 10.1016/0020-0255(86)90006-x

Krishnapuram, R., and Keller, J.-M. (1993). A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* 1, 98–110. doi: 10.1109/91.227387

Krishnapuram, R., and Keller, J.-M. (1996). The possibilistic c-means algorithm: insights and recommendations. *IEEE Trans. Fuzzy Syst.* 4, 385–393. doi: 10.1109/91.531779

Lan, Z., Sourina, O., Wang, L., Scherer, R., and Müller-Putz, G. R. (2019). Domain adaptation techniques for eeg-based emotion recognition: a comparative study on two public datasets. *IEEE Trans. Cogn. Dev. Syst.* 11, 85–94. doi: 10.1109/tcds.2018.2826840

Li, Y., Zheng, W., Cui, Z., Zhang, T., and Zong, Y. (2018). "A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition," in *27th International Joint Conference on Artificial Intelligence (IJCAI)*, New York, NY.

Li, Y.-F., and Zhou, Z.-H. (2011). "Improving semi-supervised support vector machines through unlabeled instances selection," in *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, San Francisco, CA, 386–391.

Li, Y.-F., Kwok, J.-T., and Zhou, Z.-H. (2009). "Semi-supervised learning using label mean," in *Proceedings of the 26th International Conference on Machine Learning*, Montreal, QC, 633–640. doi: 10.1145/1553374.1553456

Liu, W., Wang, J., and Chang, S.-F. (2012). Robust and scalable graph-based semi-supervised learning. *Proc. IEEE* 100, 2624–2638. doi: 10.1109/JPROC.2012.2197809

Mühl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain Comput. Interf.* 1, 66–84. doi: 10.1080/2326263x.2014.912881

Musha, T., Terasaki, Y., Haque, H. A., and Ivamitsky, G. A. (1997). Feature extraction from EEGs associated with emotions. *Artific. Life Robot.* 1, 15–19. doi: 10.1007/bf02471106

Nie, F., Xu, D., Tsang, I., and Zhang, C. (2010). Flexible manifold embedding: a framework for semi-supervised and unsupervised di- mension reduction. *IEEE Trans. Image Process.* 19, 1921–1932. doi: 10.1109/tip.2010.2044958

Simonyan, K., and Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representations (ICLR)*, New Orleans, LA, 1–14.

Song, T., Zheng, W., Song, P., and Cui, Z. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *Proc. IEEE Trans. Affect. Comput.* 11, 532–541. doi: 10.1109/taffc.2018.2817622

Tao, J. W., Song, D., Wen, S., and Hu, W. (2017). Robust multi-source adaptation visual classification using supervised low-rank representation. *Patt. Recogn.* 61, 47–65. doi: 10.1016/j.patcog.2016.07.006

Tao, J., Wen, S., and Hu, W. (2015). L1-norm locally linear representation regularization multi-source adaptation learning. *Neural Netw.* 69, 80–98. doi: 10.1016/j.neunet.2015.01.009

Tao, J., Wen, S., and Hu, W. (2016). Multi-source adaptation learning with global and local regularization by exploiting joint kernel sparse representation. *Knowl. Based Syst.* 98, 76–94. doi: 10.1016/j.knosys.2016.01.021

Tao, J., Zhou, D., Liu, F., and Zhu, B. (2019). Latent multi-feature co-regression for visual recognition by discriminatively leveraging multi-source models. *Patt. Recogn.* 87, 296–316. doi: 10.1016/j.patcog.2018.10.023

Tu, W., and Sun, S. (2013). Semi-supervised feature extraction for EEG classification. *Patt. Analys. Appl. Paa* 16, 213–222. doi: 10.1007/s10044-012-0298-2

Vapnik, V.-N. (1995). *The Nature of Statistical Learning Theory*. New York, NY: Springer, 69–83. doi: 10.1007/978-1-4757-2440-0

Wang, F., and Zhang, C. (2008). Label propagation through linear neighborhoods. *IEEE Trans. Knowl. Data Eng.* 20, 55–67. doi: 10.1109/tkde.2007.190672

Wang, Q.-W., Li, Y.-F., and Zhou, Z.-H. (2019). "Partial label learning with unlabeled data," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19)*, Macao, 3755–3761. doi: 10.24963/ijcai.2019/521

Wang, T.-Z., Huang, S.-J., and Zhou, Z.-H. (2019). "Towards identifying causal relation between instances and labels," in *Proceedings of the 19th SIAM International Conference on Data Mining (SDM'19)*, Calgary, 289–297. doi: 10.1137/1.9781611975673.33

Wang, Y.-Y., Chen, S.-C., and Zhou, Z.-H. (2012). New semi-supervised classification method based on modified cluster assumption. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 689–702. doi: 10.1109/tnnls.2012.2186825

Wang, Y.-Y., Meng, Y., Fu, Z.-Y., and Xue, H. (2017). Toward-s safe semi-supervised classification: adjusted cluster assumption via clustering. *Neural Process. Lett.* 46, 1–12.

Wu, M.-S., and Deng, X.-G. (2018). Semi-supervised pattern classification method based on Tri-DE-ELM. *Comput. Eng. Appl.* 54, 109–114.

Xue, H., and Chen, S.-C. (2007). "Alternative robust local embedding," in *Proceedings of the International Conference Wavelet Analysis Pattern Recognition*, Paris.

Xue, H., Chen, S.-C., and Yang, Q. (2011). Structural regularized support vector machine: a framework for structural large margin classifier. *IEEE Trans. Neural Netw.* 22, 573–587. doi: 10.1109/tnn.2011.2108315

Zhang, Y., Chung, F. L., and Wang, S. (2019a). Takagi-sugeno-kang fuzzy systems with dynamic rule weights. *J. Intellig. Fuzzy Syst.* 37, 8535–8550. doi: 10.3233/jifs-182561

Zhang, Y., Dong, J., Zhu, J., and Wu, C. (2019b). Common and special knowledge-driven TSK fuzzy system and its modeling and application for epileptic EEG signals recognition. *IEEE Access* 7, 127600–127614. doi: 10.1109/access.2019.2937657

Zhang, Y., Li, J., Zhou, X., Zhou, T., Zhang, M., Ren, J., et al. (2019c). A view-reduction based multi-view TSK fuzzy system and its application for textile color classification. *J. Ambient Intellig. Hum. Comput.* 1–11. doi: 10.1007/s12652-019-01495-9

Zhang, Y., Tian, F., Wu, H., Xingyun, G., Xiaofeng, Z., Lemin, T., et al. (2017). Brain MRI tissue classification based fuzzy clustering with competitive learning. *J. Med. Imag. Health Inform.* 7, 1654–1659. doi: 10.1166/jmihi.2017.2181

Zhang, Y., Wang, L., Wu, H., Xingyun, G., Yao, D., Dong, J., et al. (2016). A clustering method based on fast exemplar finding and its application on brain magnetic resonance images segmentation. *J. Med. Imag. Health Inform.* 6, 1337–1344. doi: 10.1166/jmihi.2016.1923

Zhang, Y., Chung, F., and Wang, S. (2020a). Clustering by transmission learning from data density to label manifold with statistical diffusion. *Knowl. Based Syst.* 193:105330. doi: 10.1016/j.knosys.2019.105330

Zhang, Y., Wang, S., Xia, K., Jinag, Y., Qian, P., For the Alzheimer's Disease Neuroimaging Initiative, et al. (2020b). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inform. Fusion* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002

Zhao, P., and Zhou, Z.-H. (2018). "Label distribution learning by optimal transport," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI'18)*, New Orleans, LA, 4506–4513.

Zheng, W.-L., and Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Autonom. Ment. Dev.* 7, 162–175. doi: 10.1109/tamd.2015.2431497

Zheng, W.-L., Liu, W., Lu, Y.-F., Lu, B.-L., and Cichocki, A. (2019). EmotionMeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybernet.* 49, 1110–1122. doi: 10.1109/tcyb.2018.2797176

Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* 2020:1. doi: 10.1109/taffc.2020.2994159

Zhou, D., Bousquet, O., Lal, T., Weston, J., and Schlkopf, B. (2003). "Learning with local and global consistency," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '03)*, (Vancouver: Curran Associates Inc), 321–328.

Zhou, S.-H., Liu, X.-W., Zhu, C.-Z., Liu, Q., and Yin, J. (2014). "Spectral clustering-based local and global structure preservation for feature selection," in *Proceedings of 2014 International Joint Conference on Neural Networks*, Beijing, 550–557.

Zhou, X., Jin, K., Shang, Y., and Guo, G. (2018). Visually interpretable representation learning for depression recognition from facial Im-ages. *IEEE Trans. Affect. Comput.* 11, 542–552. doi: 10.1109/TAFFC.2018.2828819

Zhou, Z.-H., and Li, M. (2010). Semi-supervised learning by disagreement. *Knowl. Inform. Syst.* 24, 415–439. doi: 10.1007/s10115-009-0209-z

Zhu, X., Ghahramani, Z., and Lafferty, J. (2003). "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proceedings of the 20th International Conference on Machine Learning*, (California: AAAI Press), 912–919.

Zhu, X.-J. (2008). *Semi-Supervised Learning Literature Survey. Computer Science TR 1530*. Madison: University of Wisconsin Madiso.

Zhu, X.-J., and Goldberg, A. (2009). *Introduction to Semi-Supervised Learning*. San Rafael, CA: Morgan & Claypool.

Zhu, Y., Shang, Y., Shao, Z., and Guo, G. (2017). "Automated depression diagnosis based on deep networks to encode facial appearance and dynamics," in *Proceedings of the IEEE Transactions on Affective Computing*, Cambridge. doi: 10.1109/TAFFC.2017.2650899

Zu, B.-B., Xia, K.-W., Wen, L., Niu, W.-J., and Jiang, X.-Q. (2019). Semi-supervised classification application of remote sensing image based on block low rank images. *J. Front. Comput. Sci. Technol.* 13, 1217–1226.

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 30 June 2021
doi: 10.3389/fnins.2021.684469

Monitoring and Prognostic Analysis of Severe Cerebrovascular Diseases Based on Multi-Scale Dynamic Brain Imaging

*Suting Zhong[1], Kai Sun[2], Xiaobing Zuo[1] and Aihong Chen[1]**

[1] Department of Emergency Medicine, Hanyang Hospital, Wuhan University of Science and Technology, Wuhan, China, [2] Department of Neurosurgery, Yantai Penglai Traditional Chinese Medicine Hospital, Yantai, China

Severe cerebrovascular disease is an acute cerebrovascular event that causes severe neurological damage in patients, and is often accompanied by severe dysfunction of multiple systems such as breathing and circulation. Patients with severe cerebrovascular disease are in critical condition, have many complications, and are prone to deterioration of neurological function. Therefore, they need closer monitoring and treatment. The treatment strategy in the acute phase directly determines the prognosis of the patient. The case of this article selected 90 patients with severe cerebrovascular disease who were hospitalized in four wards of the Department of Neurology and the Department of Critical Care Medicine in a university hospital. The included cases were in accordance with the guidelines for the prevention and treatment of cerebrovascular diseases. Patients with cerebral infarction are given routine treatments such as improving cerebral circulation, protecting nutrient brain cells, dehydration, and anti-platelet; patients with cerebral hemorrhage are treated within the corresponding safe time window. We use Statistical Product and Service Solutions (SPSS) Statistics21 software to perform statistical analysis on the results. Based on the study of the feature extraction process of convolutional neural network, according to the hierarchical principle of convolutional neural network, a backbone neural network MF (Multi-Features)—Dense Net that can realize the fusion, and extraction of multi-scale features is designed. The network combines the characteristics of densely connected network and feature pyramid network structure, and combines strong feature extraction ability, high robustness and relatively small parameter amount. An end-to-end monitoring algorithm for severe cerebrovascular diseases based on MF-Dense Net is proposed. In the experiment, the algorithm showed high monitoring accuracy, and at the same time reached the speed of real-time monitoring on the experimental platform. An improved spatial pyramid pooling structure is designed to strengthen the network's ability to merge and extract local features at the same level and at multiple scales, which can further improve the accuracy of algorithm monitoring by paying a small amount of additional computational cost. At the same time, a method is designed to strengthen the use of low-level features by improving the network structure, which improves the algorithm's monitoring

OPEN ACCESS

Edited by:
Yizhang Jiang, Jiangnan University, China

Reviewed by:
Jing Xue, Wuxi People's Hospital Affiliated to Nanjing Medical University, China
Xiaoqing Gu, Changzhou University, China

***Correspondence:**
Aihong Chen
chenaihong16@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *23 March 2021*
Accepted: *26 April 2021*
Published: *30 June 2021*

Citation:
Zhong S, Sun K, Zuo X and Chen A (2021) Monitoring and Prognostic Analysis of Severe Cerebrovascular Diseases Based on Multi-Scale Dynamic Brain Imaging. Front. Neurosci. 15:684469. doi: 10.3389/fnins.2021.684469

performance on small-scale severe cerebrovascular diseases. For patients with severe cerebrovascular disease in general, APACHEII1, APACHEII2, APACHEII3 and the trend of APACHEII score change are divided into high-risk group and low-risk group. The overall severe cerebrovascular disease, severe cerebral hemorrhage and severe cerebral infarction are analyzed, respectively. The differences are statistically significant.

Keywords: severe cerebrovascular disease, dynamic brain imaging, multi-scale features, ApacheII score, monitoring and prognostic analysis

INTRODUCTION

Severe cerebrovascular disease is a critical type of acute cerebrovascular disease, which causes severe neurological damage in patients, and is often accompanied by severe dysfunction of multiple systems such as breathing and circulation (Conzen et al., 2019). Because patients with severe cerebrovascular disease are critically ill, have many complications, and are prone to deterioration of neurological function, they often need to be transferred to the intensive care unit for closer monitoring and treatment (Zafar et al., 2019; Xu et al., 2020). The activation of the neuroendocrine system, the increase in intracranial pressure, and a history of previous hypertension after acute stroke often lead to increased blood pressure (Václavů et al., 2019). Therefore, blood pressure management is one of the basic links and one of the crucial links in the acute treatment of patients with severe cerebrovascular disease (Rosenberg et al., 2020). Elevated blood pressure can lead to encephalopathy, cardiac complications, and kidney damage. In patients with spontaneous cerebral hemorrhage, it can also cause hematoma enlargement, aggravate the edema of the tissue around the hematoma, and cause rebleeding (Savić et al., 2019). However, excessive reduction in blood pressure can increase the degree of cerebral ischemia in patients with acute ischemic stroke by reducing cerebral perfusion pressure, and will also significantly increase the risk of hematoma enlargement, rebleeding, death, and disability in patients with spontaneous cerebral hemorrhage (Rocha and Singhal, 2020).

In the neurological intensive care unit, the second injury in the course of most patients with acute brain injury will significantly increase the rate of deterioration and mortality (Robertson et al., 2019). Early detection and early prevention of these complications are necessary. Before the general neurological examination is positive, the brain function or structure has undergone irreversible changes. The brain function monitoring will provide effective information during the reversible period of neurological dysfunction, early detection of the second injury, help clinicians to intervene in time to prevent continuous brain damage, and continuous evaluation of the treatment effect (Mathieu et al., 2020). EEG (Electroencephalogram) is more sensitive to abnormal brain metabolism, ischemia, hypoxia and abnormal nerve function, and can be used as one of the methods to monitor brain function in the NICU (Neonatal Intensive Care Unit) (Merikangas et al., 2019). It can determine non-convulsive seizures (NCS), non-convulsive status epilepticus (NCSE), guide the drug treatment of epilepsy, predict and detect cerebral ischemia, monitor cerebral edema and intracranial pressure, and help predict the outcome of coma patients (Quon et al., 2019). EEG is closely related to the metabolism of brain nerve cells. It is sensitive to brain damage caused by ischemia and hypoxia. Nervous system dysfunction can be found in the reversible phase of brain damage, and the recovery of nerve function can be found when other clinical examinations have failed (Rahmim et al., 2019). It is the most effective means to discover epileptic activity. In addition, CEEG (Clinical Electroencephalogram) monitoring can provide dynamic information about brain function, which is conducive to real-time monitoring of brain function changes. The rapid development of computer and computer network technology makes the implementation and wide application of CEEG monitoring possible (Tanaka et al., 2020). Digital EEG (DEEG) has become the preferred technology of CEEG. It makes it easy to store a large amount of EEG information, which can be edited, quantitatively analyzed, automatically removed artifacts, and automatically identified abnormal waveforms (Marini et al., 2020). CEEG accurately arranges disk electrodes in accordance with international standards. Monitoring requires good scalp electrode fixation. Generally, electrode paste is adhered to the scalp surface and fixed with collodion, which is feasible for long-term monitoring (Lohith et al., 2019; Malas et al., 2019). The standard EEG recorder records CEEG in the storage medium in digital format. In NICU, DEEG sampling frequency of 128–256/channel/s can provide enough information. The detailed EEG data of each patient every day can be stored up to 2GB. After obtaining the EEG data, the NICU staff needs to review and summarize, and the EEG technicians and EEG experts will observe and analyze, and draw conclusions (Taneja et al., 2019; Maïer et al., 2020). Although NICU nurses can recognize some typical EEG after training, some complicated CEEG records still require detailed explanation by EEG experts (Waddle et al., 2020). This has always been a problem that restricts the widespread use of CEEG for NICU monitoring and needs to be further studied. The original EEG is converted into QEEG (Quantified Electroencephalogram). It mainly includes power spectrum analysis, time domain analysis, formation of compressed spectrum array and EEG topographic map (Khalil et al., 2020). QEEG can quickly display important EEG changes, such as epileptic waves, slow wave, general suppression, lack of fast wave, increase and decrease of EEG variability, etc. But at present, QEEG still cannot be used in clinic separately from the original EEG (Koskinen et al., 2019).

Nerve cells are very sensitive to ischemic damage, and the pyramidal cells of the third and fifth layers of the cerebral cortex, which are closely related to the production of EEG, are most vulnerable to ischemic damage (Wang C. et al., 2019). Ischemia

damage causes changes in the posterior potential of cerebral cortex neurons, and abnormal information can be recorded in the scalp EEG. When the cerebral blood flow (CBF) drops to 25–30 ml/100 g/min, the neuronal dysfunction is in a reversible phase, and the EEG changes (Wang Y. et al., 2019). It is often manifested as a lack of α and β activities, and theta waves and delta waves appear one after another. When CBF drops below 10–12 ml/100 g/min, nerve cells will be irreversibly damaged, and EEG will also be abnormal (Li et al., 2020). Therefore, EEG is very sensitive to cerebral ischemia during the reversible period of neuronal dysfunction, and there are general changes. Clinical timely intervention can prevent continuous brain damage. On the other hand, EEG is also very sensitive to the recovery of cerebral ischemia, and it is found earlier than clinical examination. Cerebral edema and increased intracranial pressure in the course of cerebral hemorrhage patients are the main reasons for the deterioration and death of the disease. Monitoring it is the key to prevention and treatment (Wardlaw et al., 2019). In patients with cerebral edema and increased intracranial pressure, EEG shows continuous diffuse wave activity, and EEG is significantly improved after using dehydrating agents such as mannitol (Zeiler et al., 2020). CEEG monitoring can not only be used to monitor and reflect dehydration of dehydrating agents in treatment of cerebral edema, but also can provide early effects of drug treatment. Compared with CT, MRI, and other examinations, CEEG has the advantage of continuous, dynamic and real-time bedside monitoring (Liu et al., 2019). Transcranial Doppler (TCD) ultrasound can evaluate the cerebral perfusion and intracranial pressure of patients with severe cerebrovascular disease by monitoring the changes of hemodynamic parameters, such as blood flow direction, blood flow velocity, spectrum shape, pulsation index, etc. After severe cerebrovascular disease occurs, intracranial pressure changes sharply, and brain herniation often occurs (Forti et al., 2019; Dumitrascu and Koronyo-Hamaoui, 2020). The initial manifestation of TCD changes when the intracranial pressure increases, the systolic peak becomes sharp, and the diastolic blood flow signal disappears when the intracranial pressure approaches the diastolic pressure, leaving only the systolic sharp wave; the intracranial pressure continues to increase, when the diastolic blood pressure is exceeded, the diastolic blood flow reappears, but the direction is negative, which is a "shock wave," indicates that the compensation mechanism of cerebrovascular is not enough to counter the increase in intracranial pressure. It shows needle-like blood flow in the early stage of contraction, that is, "nail wave." At this time, it is difficult for peripheral blood flow to enter the cerebral circulation, which is a characteristic manifestation of cerebral circulation stop (Dehkharghani and Qiu, 2020). The flow is getting smaller and smaller, and eventually the blood flow stops.

In this experimental study, the EEG patterns of severe brain patients were also analyzed and studied, and it was pointed out that the prognosis of patients with a-wave coma was relatively poor. In severe patients without a wake-sleep cycle, if some sleep spindle components are added to the slow wave or a wave background, it indicates that the patient has a better prognosis, especially the long-term EEG monitoring in the diagnosis, treatment and evaluation of acute and severe cerebrovascular diseases. Aiming at the monitoring problem of severe cerebrovascular disease in multi-scale dynamic brain imaging, the bottleneck of the existing convolutional neural network for monitoring severe cerebrovascular disease in the multi-scale feature extraction is analyzed, and the dense connection network structure and the feature pyramid network structure are combined to design a convolutional neural network MF-Desne Net that can realize multi-scale feature fusion extraction. Based on this network, an end-to-end multi-scale dynamic brain imaging monitoring algorithm for severe cerebrovascular disease MDRD (Multi-scale Dense ResNet Detector) was designed. Experiments show that the MDRD algorithm not only achieves high monitoring accuracy, but also has a fast monitoring speed, which can meet the needs of real-time monitoring. At the same time, on the basis of the above work, the multi-scale feature extraction of convolutional neural network is further studied. A three-dimensional spatial pyramid pooling method is proposed to improve the network's ability to fuse and extract local multi-scale features of the image; a method to strengthen the use of low-level features based on reconstruction operations is proposed, which improves the network's ability to use low-level extraction capacity. APACHEII1, APACHEII2, and APACHEII3 are related to prognosis. The higher the score, the higher the risk of death. The APACHEII1, APACHEII2, and APACHEII3 of the death group are higher than those of the survival group. The scores of APACHEII1, APACHEII2, and APACHEII3 according to Youden index are 19, 19, and 17 points, respectively. It shows that the score of patients with severe cerebrovascular disease is higher than 19 in the first 24 h after admission to the NICU, the score is still higher than 19 in the second 24 h, and the score does not fall below 17 in the third 24 h.

The rest of this article is organized as follows. Section "Data and Methods for Monitoring Severe Cerebrovascular Disease" discusses the data and methods of monitoring severe cerebrovascular diseases. In section "Multi-Scale Dynamic Brain Imaging Monitoring Model for Severe Cerebrovascular Disease Based on Feature Fusion", a multi-scale dynamic brain imaging monitoring model of severe cerebrovascular disease based on feature fusion is constructed. In section "Analysis of the Results of Monitoring and Prognosis of Severe Cerebrovascular Disease", the results of the surveillance and prognosis experiments of severe cerebrovascular disease were analyzed. Section "Conclusion" summarizes the full text.

DATA AND METHODS FOR MONITORING SEVERE CEREBROVASCULAR DISEASE

Clinical Features of Severe Cerebrovascular Disease

In clinical practice, patients with severe cerebrovascular diseases have varying degrees of consciousness disturbance, and the prognosis is poor. Patients with severe cerebrovascular diseases are often prone to hospital-acquired pneumonia due to the age of the patient, poor patient awareness, poor cough and sputum

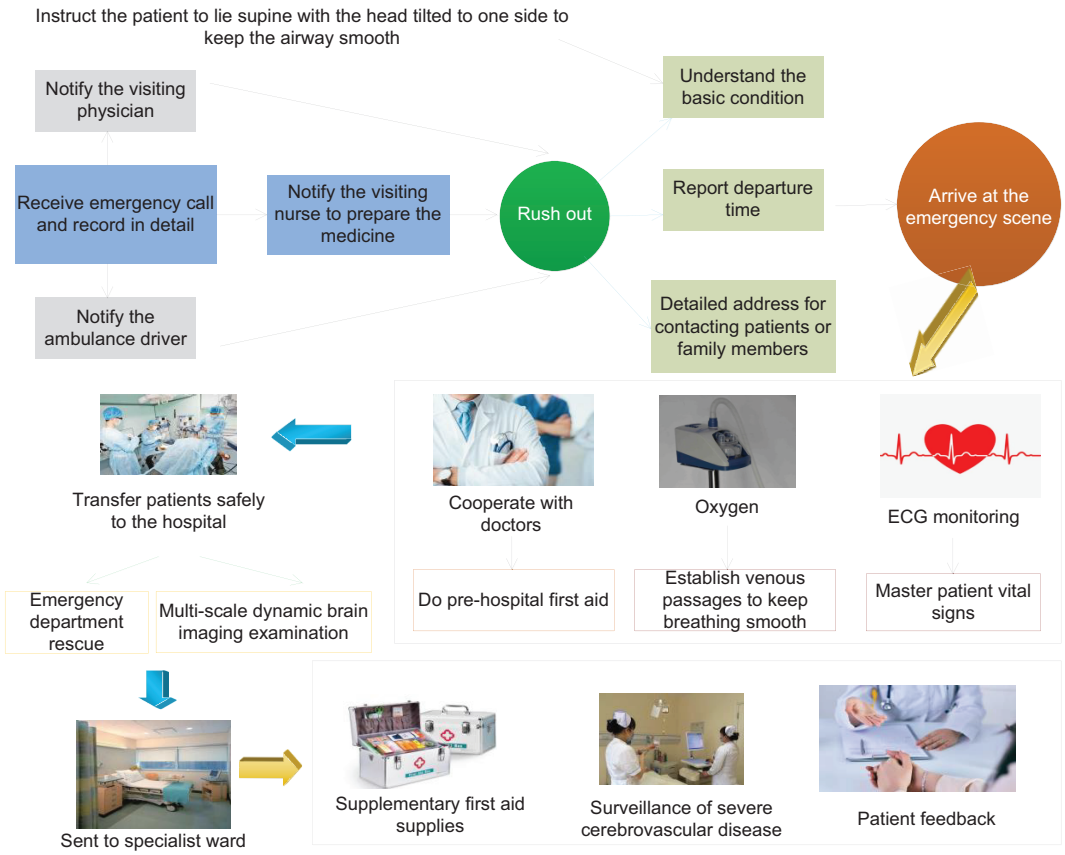


FIGURE 1 | Flow chart of emergency care for patients with severe cerebrovascular disease.

expectoration, and the use of multiple medical equipment, which will seriously affect the prognosis of the patient. In addition, it is neurogenic. Pulmonary edema can also aggravate the progression of pneumonia. Severe cerebrovascular disease is accompanied by multiple complications, such as pneumonia, electrolyte imbalance and concurrent seizures, and it is found that patients with concurrent pulmonary infection and internal environmental disorders indicate a higher mortality rate. Half of the patients included in the study also developed pneumonia to varying degrees, and the patients developed severe pneumonia and were infected with multi-drug resistant bacteria. The patient underwent tracheotomy, confirming that pneumonia is a common complication of severe cerebrovascular disease.

More than half of the surviving patients with severe cerebrovascular disease will have different degrees of disability, such as paralysis of the lateral limbs, various types of aphasia, etc. This type of disease brings heavy economic pressure to the patient's family and the entire society. The most common cause of stroke is cerebrovascular disease, of which hypertensive arteriosclerosis and atherosclerosis lead to vascular disease. The overall prognosis of patients with cerebrovascular disease with mild consciousness disorder is relatively good, and the survival rate is higher; the prognosis of patients with moderate consciousness disorder is poor, and most of them will have different degrees of sequelae. Patients with severe disturbance of consciousness have the highest mortality rate and the worst prognosis. Surviving patients have serious sequelae, which seriously affect the quality of life of patients. **Figure 1** shows the flow chart of emergency care for patients with severe cerebrovascular disease.

Patients with severe cerebrovascular disease often have varying degrees of consciousness disturbance. In clinical practice, consciousness refers to the degree of awakening of the brain, which is the ability of the cerebral nervous system to respond to various internal and external stimuli in a timely manner. The subject of Consciousness Disorder is divided into the decline in the degree of arousal and the changes in the content of consciousness. The content of consciousness refers to the high-level neural activities of the cerebral cortex, including the mental activities of human perception, thinking, memory, orientation, emotion, and will activity. Consciousness content is manifested as confusion or delirium. The brainstem ascending activation system plays an important role in the maintenance of human consciousness. It receives various sensory signals and transmits them to the non-specific nucleus of the thalamus, and then projects to the extensive cerebral cortex, making the central nervous system in a state of excitement. Once a patient develops a disturbance of consciousness, indicating a critical condition, the

prognosis of the patient is often poor. In clinical work, being able to recognize various types of consciousness disorders in patients in a timely and accurate manner, and be able to fully evaluate the patient's condition and make correct treatments can enable patients with severe cerebrovascular disease to be rescued in time, and even turn the crisis into peace.

EEG Physiological Monitoring for Brain Function Assessment

The cerebral cortex cells have spontaneous electrophysiological activities. When the brain is damaged by various reasons, it will show abnormal discharges, including changes in frequency and amplitude. EEG technology is to record these brain electrical activities through brain electrodes, and through the processing of computer technology, these brain electrical activities are turned into readable graphics that can more accurately reflect brain electrical activities. The potential on the EEG reflects the postsynaptic potential of the gray matter vertebral body cells, and the rhythm of the EEG is maintained by the brainstem network ascending activation system.

Electroencephalography technology uses sophisticated instruments to measure the rhythmic and spontaneous bioelectric activity of the cerebral cortex, and amplifies the recorded bioelectric signals by millions of times to understand the functional state of the brain. It is often used in the diagnosis and treatment of epilepsy in clinical practice. It is an important method for the diagnosis and classification of epilepsy. It can conveniently, objectively, safely, non-invasively and accurately evaluate the brain function of patients.

The placement of the EEG electrodes adopts the 10–20 system electrode placement method recommended by the International Electroencephalography Society. Its characteristic is that the arrangement of each brain electrode is proportional to the size and shape of the head, and the electrode name matches the brain anatomy zone. With the apex as the center of the circle, you draw straight lines to the temporal halves (divided into 10), and then make concentric circles with the halves of the sagittal line as the radius, and determine the electrode placement position according to the intersection point.

The condition of EEG is closely related to the metabolism of brain cells, and it is more sensitive to brain tissue damage caused by various causes. When brain cells undergo hypoxia and other pathological changes such as brain cell edema and necrosis, the EEG will show amplitude. The frequency changes, and the location of the lesion can be roughly located according to the position of the electrode, which plays an important role in the basic diagnosis process. For patients with severe cerebrovascular disease in clinic, an ordinary and short EEG cannot better reflect the patient's brain function status. Long-term EEG can monitor the EEG physiological activities of severe patients for a long time at the bedside and provide long-term dynamic information.

Clinical Analysis of Long-Term EEG

Severe cerebrovascular diseases are often caused by important responsible vascular lesions. In patients with common large-area cerebral infarctions, due to ischemia, edema, and necrosis of brain cells in the infarct focus, the brain cell potential in this part will change. Generally, the brain cells in the central area of the infarct focus have been necrotic, and the EEG generally shows a flat waveform with low amplitude. The brain cells adjacent to the peripheral edema zone of the lesion will also experience varying degrees of hypoxia. Generally, the EEG shows a low amplitude slow wave (θ wave or δ wave). Clinically, the severity of the patient's condition can be judged based on the flat and slow wave range and amplitude.

In the early stage of the onset, some patients with cerebrovascular accidents, especially acute cerebral ischemic diseases, cannot find positive signs in time by neuroimaging examination alone. EEG is closely related to brain bio-metabolism and more accurately reflects the cerebral cortex. Brain bioelectrical activity can provide timely and accurate information on the state of brain functions closely related to abnormal brain metabolism. EEG is a more sensitive indicator for monitoring acute ischemic diseases. Infarction has important clinical diagnostic value and is worthy of promotion and application. When the local CBF is reduced to 20–30 ml/100 g/min, the EEG monitoring will appear abnormal. High amplitude waves can be seen to replace the background waves of the EEG. When the blood flow is reduced to 17 ml/(100 g min), synaptic activity will drop and the wave will disappear. When the blood flow is reduced to 10 ml/100 g/min, brain cells will be irreversibly damaged. Continuous EEG monitoring is of great significance for the early diagnosis of patients with acute vascular accidents. As patients with severe cerebrovascular diseases will have varying degrees of consciousness disturbances, physical examinations are poorly coordinated, and it is more subjective to judge the severity of the disease through neurological examinations, and EEG monitoring is extremely important. The human body intention perception based on brain muscle multi-source perception needs to judge the movement intention of brain electrical signals, so as to realize the active and passive coordinated control of brain muscle signals, as shown in **Figure 2**.

Abnormal EEG and brain topography show that the absolute power value of the fast-wave frequency band decreases, and the absolute power value of the slow-wave frequency band increases. Therefore, the EEG examination can more sensitively reflect the changes in brain function of patients before cerebral hemorrhage. The reason may be that brain damage has occurred in the early stage of hypertensive intracerebral hemorrhage. This examination provides reference and basis for early prediction and preventive treatment of the occurrence of intracerebral hemorrhage. Therefore, long-term EEG monitoring for patients with hypertension, especially the extremely high-risk group, is of clinical significance for the timely detection of cerebral hemorrhage.

The brain damage caused by non-convulsive epilepsy and NCSE and the damage of primary cerebrovascular disease are not only simply superimposed, but can make the condition further worse. After initial treatment of convulsive status epilepticus, if the interictal EEG shows epileptic discharges, periodic discharges, and persistent micro-seizure patterns, the risk of recurrence of convulsive status epilepticus increases by 5,

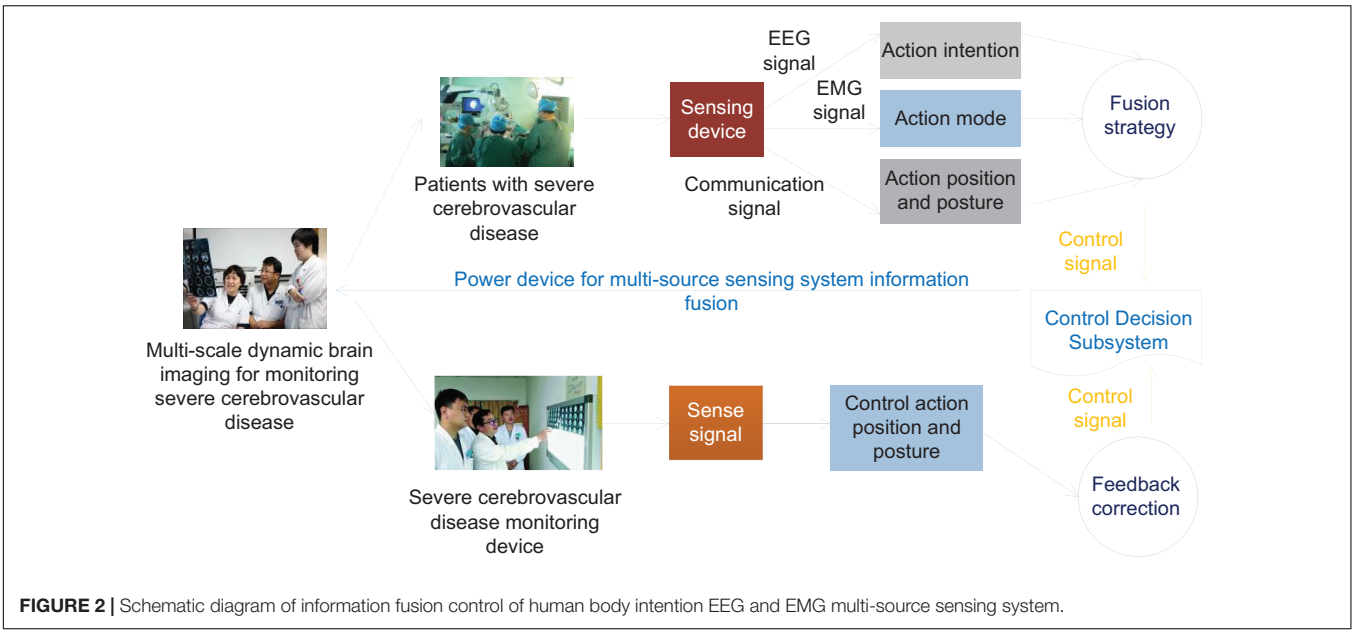


FIGURE 2 | Schematic diagram of information fusion control of human body intention EEG and EMG multi-source sensing system.

18, and 18, respectively. As we all know, electroencephalogram is an important auxiliary examination for the diagnosis of epilepsy, and it will not be described here. However, due to conditions and time constraints, the positive rate of diagnosis of epilepsy is not high. Some patients with severe encephalopathy have NCS in clinical practice. If there is no long-term EEG monitoring, it is difficult to find non-convulsive epilepsy in clinical work. Among the 90 patients in this study, there were 24 cases of epilepsy, and the incidence rate was 26.67%. However, only six cases had spastic seizures, the positive rate was only 6.6%, and the incidence of non-convulsive epilepsy was 20%, so non-convulsive epilepsy occurred in 20%. The incidence of convulsive epilepsy in patients with severe cerebrovascular disease is much higher than that of convulsive epilepsy. Long-term EEG is an important means to find non-convulsive epilepsy.

Long-term EEG monitoring can also be applied to the monitoring of epilepsy drug treatment. Insufficient dose of antiepileptic drugs cannot control epilepsy well. Excessive medication will increase the patient's respiratory and cardiovascular diseases. Therefore, long-term EEG monitoring is extremely important.

Research Objects for Surveillance of Severe Cerebrovascular Disease

The cases were selected from 90 patients with severe cerebrovascular disease in the Department of Neurology and Critical Care Medicine of the affiliated hospital of a medical university, including 42 male patients and 48 female patients, including 34 patients with cerebral hemorrhage and 56 patients with cerebral infarction. The exclusion criteria are shown in **Table 1**.

Telephone follow-up was uniformly used for 90 days. Among them, the patients who died within 90 days completed the follow-up, and the rest continued to follow-up until the patient was discharged 3 months later to understand the prognosis. Statistics of general conditions include age, gender, hypertension, blood lipids, type of cerebrovascular disease, GCS (Glasgow Coma Scale) score, EEG classification and typical EEG, etc., to analyze their relationship with the patient's prognosis.

Patients with large-area cerebral infarction are limited to hospital emergency diagnosis by brain imaging examination as internal carotid artery or middle cerebral artery main trunk occlusion and unclassified cerebral vascular infarction. The cerebral infarction area exceeds a single lobe and the lesion area is more than 5 cm.

TABLE 1 | Exclusion criteria.

Exclusion criteria number	Exclusion criteria description
1	Does not meet the basic standards
2	The patient's family members ask to give up treatment after admission
3	Rectal body temperature < 32°C
4	Use drugs that significantly affect the determination of brain electrical activity before the experiment
5	Combined with serious diseases or factors that affect the judgment of brain function (such as hepatic encephalopathy, pulmonary encephalopathy, renal encephalopathy, toxic encephalopathy, autoimmune diseases, infectious diseases, brain tumors, etc.)
6	Nervous system examination is highly suspected of cerebral infarction but not confirmed by brain MRI
7	The patient does not breathe spontaneously and requires a ventilator to assist breathing
8	The patient's limb jitter affects EEG acquisition
9	Past history of mental disorders

Intracerebral hemorrhage patients were admitted to the hospital by CT craniocerebral hemorrhage (mainly near the basal ganglia).

Based on brain imaging data, the computer calculates the area of the infarct lesions at each level (in cm^2), multiplying the thickness of the septal layer of the image data (in cm) to calculate the volume of the infarct lesion. The brain blood volume of the patient is calculated according to the head MRI or CT, according to the maximum hematoma volume formula proposed by Tada's.

Research Methods for Monitoring Severe Cerebrovascular Disease

After admission, blood pressure was monitored repeatedly (measured by mercury sphygmomanometer). According to the diagnostic criteria for hypertension, patients with systolic blood pressure higher than 140 mmHg and diastolic blood pressure higher than 90 mmHg were defined as hypertensive patients.

According to the latest "Guidelines for the Prevention of Dyslipidemia in Adults in China," each patient's blood lipid level is divided: all blood samples were collected on an empty stomach, and the cholesterol (TC, TotalCholesterol) was 5.18–6.22 mmol/L for marginal increase, and cholesterol (TC) d is higher than 6.22 mmol/L; triglycerides (TG, TriGlycerides) higher than or equal to 2.26 mmol/L is higher. High density lipoprotein is (HDL, High Density Lipoprotein) $<$ 1.04 mmol/L; low density lipoprotein is (LDL, Low Density Lipoprotein) $\geq$ 4.14 mmol/L.

All cases used the EEG monitoring system (SP3000) tendered from Shanghai Shijia Medical Equipment Co., Ltd. for continuous bedside EEG monitoring, and L-EEG (Low Electroencephalogram) tracing used FP1, FP2, C3, C4, T3, T4, O1, O2. The lead electrode is connected, and the lead is placed according to the international 10/20 system. The silver plate electrode is uniformly used. After the head is cleaned, the surface is degreased with an alcohol cotton ball, the electrode is glued with conductive paste, and then fixed with a medical mesh cap. We set the paper speed to 3 cm/s, the filter channel to 0.5–35 Hz, and record 12–24 h, and save for reading.

The included cases are in accordance with the Chinese guidelines for the prevention and treatment of cerebrovascular diseases, and try to maintain consistent treatment methods without affecting the efficacy. Patients with cerebral infarction are treated with conventional antiplatelet, improvement of cerebral circulation, lipid reduction, protection of nutrient brain cells, and dehydration; patients with intracerebral hemorrhage will undergo minimally invasive removal of intracerebral hematoma within the corresponding safe time window, followed by routine irrigation and drainage to the removal of needles.

Strictly follow the inclusion and exclusion criteria of the research objects, we communicate with the EEG physicians, and pay close attention to the EEG conditions of the patients during treatment to obtain comprehensive and reliable information. The discharged patients are followed up regularly every month according to the discharge time, and the recovery situation is investigated in detail according to the rehabilitation. Data collation and data processing are strictly proofread to prevent data errors from affecting the experimental results.

MULTI-SCALE DYNAMIC BRAIN IMAGING MONITORING MODEL FOR SEVERE CEREBROVASCULAR DISEASE BASED ON FEATURE FUSION

Densely Connected Backbone Network

Res Net and VGGNet (Visual Geometry Group Network) are currently popular backbone network types in the field of severe cerebrovascular disease monitoring, but the types of networks are difficult to meet our requirements for backbone network performance. Among them, due to the limitation of the number of network layers, VGGNet has relatively weak feature extraction capabilities and a large amount of network parameters; while the deep Res Net exhibits excellent feature extraction capabilities, it also has a large network scale and computational complexity. It is difficult to meet the requirements of real-time monitoring.

In the traditional convolutional neural network design philosophy, deepening the network depth is an effective method to improve the network feature extraction ability, but as the convolutional neural network deepens, the probability of information loss after passing through the multilayer network will also increase. The network is difficult to train. This is also the main reason for limiting the depth of traditional networks such as VGGNet. Res Net uses Skip Connection to establish a direct connection path from the low-level to the high-level of the network to improve this problem and increase the network depth. However, research has shown that in the process of training the residual network, if some layers are randomly discarded, the generalization ability of the network can be improved. This shows that the residual network structure has a certain degree of redundancy. Each layer in the network only learns few features, so fewer parameters can be used for learning in each layer. This also points out that a strict progressive hierarchical structure is not necessary for neural networks. Each layer in the network does not completely depend on the features of the immediately preceding layer, and it can also be learned from the features obtained by the previous layer. Dense Net (Densely Connected Convolutional Networks) is based on the above research.

Through its Dense Connectivity (Dense Connectivity) structure brings the characteristics of feature reuse, at the same depth, the number of network parameters and the amount of calculation are greatly reduced compared with other types of convolutional neural networks. The low-level information of the network using this structure can be effectively transmitted to the high-level, and the network depth can be deepened. Dense connections and residual connections have similarities, but there are also essential differences. The difference between the feature transfer of the residual network and the densely connected network will be explained below. Suppose there is an image y0 with an L-layer convolutional neural network. Each layer of the image must undergo a non-linear conversion operation Hl(.), Hl(.) is the representative function of the network operation set of this layer, where l refers to the current number of network layers, and the output of the first layer is defined as yl, then the

feature transfer of the traditional feedforward convolutional network can be expressed by the following formula:

$$y_l = H_l \cdot (y_{l-1}) \tag{1}$$

After adding the residual connection in Res Net, it can be expressed by the following formula:

$$y_l = H_l \cdot (y_{l-1}) + y_{l-1} \tag{2}$$

In Dense Net, dense connections are used to further improve the transfer of information between layers. When the network joins the dense connection, the output of layer l can be expressed by the following formula:

$$y_l = H_l \cdot [y_0, y_1, ..., y_{l-1}] \tag{3}$$

Residual connection is to add the feature maps of the previous layer and the feature maps of the latter layer together, while dense connection is to splice the feature maps of the two together in a cascading manner, completely retaining the features output by the previous node.

The dense connection block is a special module in Dense Net. **Figure 3** shows the internal structure of a typical densely connected block. In Dense Block, a basic structure is formed by a combination of several basic network operators such as convolution operations and activation functions. Each basic structure is fed forward through dense connections. The form of connection is connected to every other basic structure. This feature makes each convolutional layer in the densely connected block only need to use a small number of convolution kernels to learn a small number of features, and the final layer will summarize and output the features learned by each layer of the network in the block. This feature allows Dense Net to learn rich features with a small number of parameters and calculations. Therefore, we can add densely connected blocks to the design of the backbone network to achieve the need to reduce the number of parameters.

In order to avoid the inconsistency of the feature map size during splicing, a densely connected network usually consists of several densely connected blocks. The size of the feature map output by each layer within each densely connected block is the same. Downsampling is performed to change the size of the feature map. **Figure 4** shows the structure of a typical densely connected network.

Backbone Network Design

In view of the advantages of the densely connected network, the network has excellent feature extraction capabilities while the network parameters and calculations are small. This paper uses the design of dense connection blocks in the basic architecture of the algorithm backbone network, and builds a fast dense connection network F-Dense Net for the monitoring task of severe cerebrovascular disease in dynamic brain imaging in this paper. The main structure of F-Dense Net includes a simplified Stem block, four densely connected blocks and three conversion layers. In ordinary Dense Net, after the input layer, a convolutional layer with a convolution kernel size of 7 and a step size of 2 and a pooling kernel size of 3 and a maximum value pooling with a step size of 2 are connected. The structural Stem block can replace the initial two layers in the ordinary Desne Net, thereby reducing the loss of the original information of the input image. Compared with the structure of ordinary Dense Net, although the Stem block retains more original information, it will introduce a lot of extra calculations. Through further analysis, we combined a 7×7 cross-convolution with a step size of 2 in the traditional Desne Net and a 2×2 maximum pooling with a step size of 2 in the Stem block. This structure can calculate and reduce the loss of original information compared with the traditional structure.

We add a conversion layer between every two densely connected blocks instead of simply pooling down sampling. The conversion layer consists of a 1×1 convolutional layer and a 2×2 maximum pooling. Among them, the number of channels of the 1×1 convolutional layer is lower than the dimension of the feature map output by the previous densely connected block. This is to fuse the features of different channels while compressing network parameters. Maximum pooling is used to down-sample the feature map, and at the same time, it can better retain some local pattern features than average pooling.

In the traditional densely connected network, Re LU (Rectified Linear Unit) is used as the activation function. Re LU has many advantages over traditional activation functions such as Sigmoid function and Tanh function. However, the Re LU activation function itself has another problem, that is, the problem of dying Re LUs. In a network that uses Re LU as the activation function, suppose that some neural nodes learn a relatively large negative deviation term in the back propagation process, so that their input becomes less than or equal to 0, then Re LU will always be in a suppressed state. Their gradient is always 0, and the network cannot learn the features on these nodes. When there are many such nodes, many nodes in the network cannot learn effectively, and the actual feature extraction ability of the network is lower than the theoretical feature expression ability of the network. The Scaled Exponential Linear Units (SELU) activation function is an activation function with self-stabilizing characteristics. The formula of SELU is:

$$selu(y) = \lambda \cdot \begin{cases} y & y > 0 \\ a \cdot e^y - a & y \leq 0 \end{cases} \tag{4}$$

The self-stabilization characteristic of SELU has an inhibitory effect on the gradient dispersion and gradient disappearance of the network. At the same time, the gradient of SELU is not 0 when the input is equal to 0, which avoids the node death problem of Re LU. In our research, we found that using SELU as the activation function in F-Dense Net has better performance than activation functions such as Re LU and Leaky Re LU, so we chose SELU as the activation function of F-Dense Net. In general, SELU has the following advantages: SELU suppresses the problems of gradient disappearance and gradient dispersion, making it very suitable for deep convolutional neural networks; SELU also avoids the dead zone problem of Re LU, and can give full play to the advantages of deep convolutional neural networks.

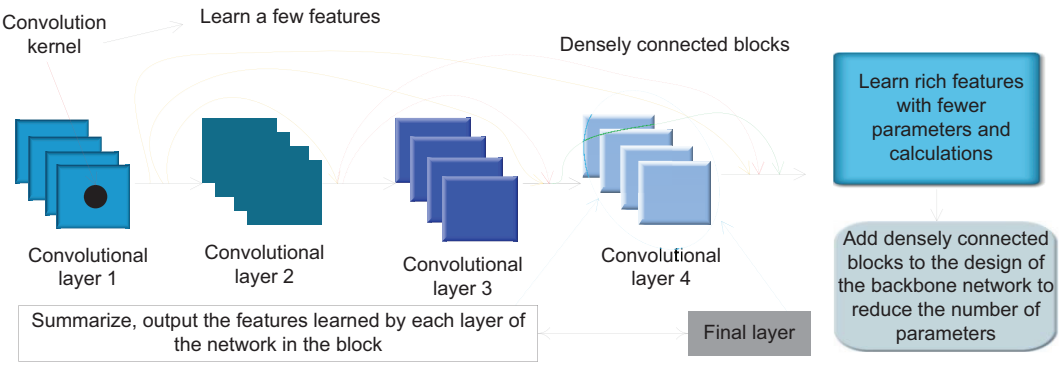


FIGURE 3 | Schematic diagram of dense connection block structure.

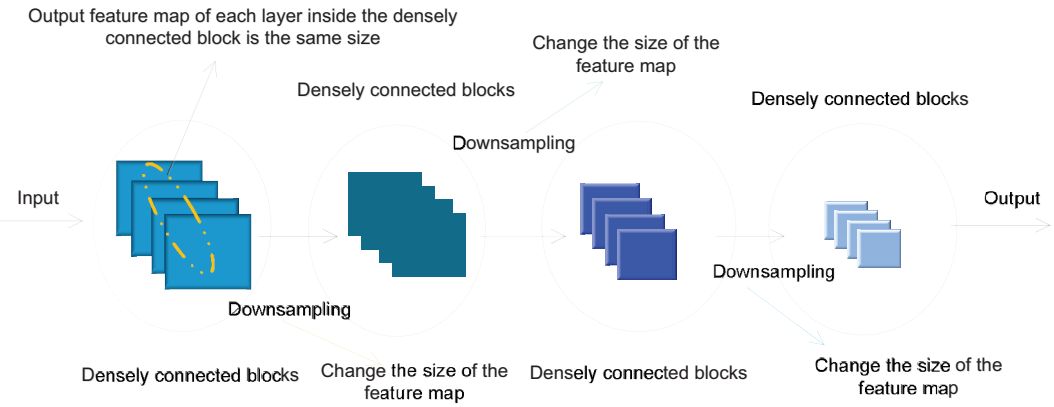


FIGURE 4 | Densely connected network structure.

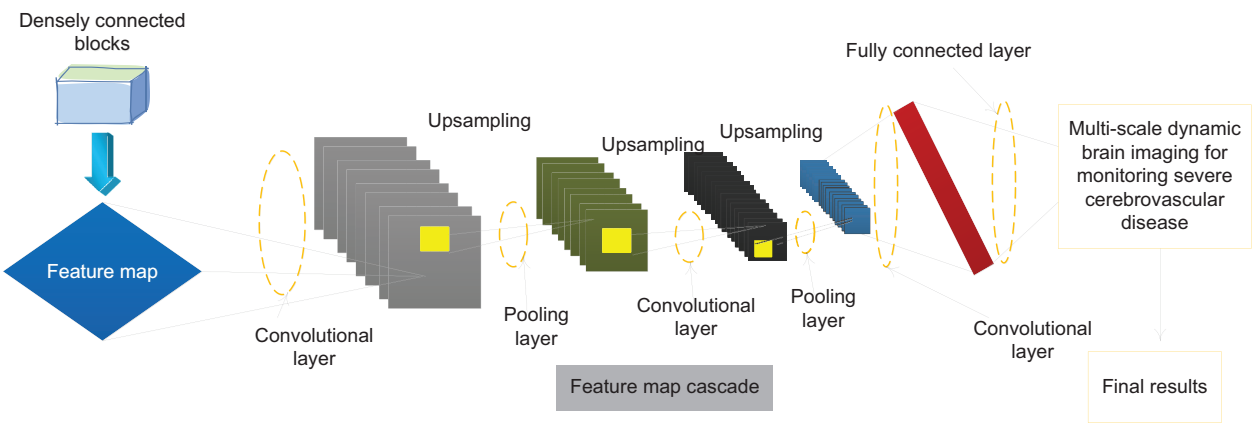


FIGURE 5 | Implementation method of feature pyramid based on MF-Dense Net.

Multi-Scale Network Structure

The feature image pyramid method improves the monitoring accuracy of severe cerebrovascular diseases at different scales by calculating features on each independent image scale. Therefore, this method introduces a lot of repetitive calculations in the training process, which brings a lot of extra time and computational resource overhead. If the feature image pyramid is used only during testing, it will lead to inconsistency between testing and training. The pyramid feature hierarchical structure method extracts features of different scales from different layers of the network for prediction, and does not bring additional calculations. However, the pyramid feature hierarchy method

does not realize the fusion of multi-level features. The lack of high-level features on the feature map for small-scale severe cerebrovascular disease monitoring makes the semantic information insufficient and prone to misdetection of small-scale severe cerebrovascular disease.

This paper adopts a new multi-scale feature extraction method, Feature Pyramid Networks (FPN) as a method to achieve multi-scale feature extraction. The FPN structure only brings a small amount of additional computational overhead on the premise of multi-level feature fusion. We improve the network architecture of F-Dense Net through the FPN structure to obtain a convolutional neural network with multi-scale feature expression capabilities. The FPN structure mainly consists of three parts: bottom-up path, top-down path and horizontal connection. The network that joins the FPN structure can solve the problem of multi-level feature fusion and extraction through the combination of these three feature transmission paths. Through the horizontal connection, the network can merge the high-level feature map with less low-level feature information and more semantic information in the top-down path and the low-level feature map with more low-level feature information and less semantic information in the bottom-up path, and this structure also provides the network with the ability to predict on multiple scale feature maps.

This paper introduces FPN on the basis of F-Dense Net to improve the network structure, and designs a network structure that supports three-scale prediction. The specific structure is shown in **Figure 5**. The bottom-up path is the forward propagation process of F-Dense Net, and the top-down path is through the last layer of F-Dense Net. The semantic information output is the richest. High-level feature maps are sampled and then followed by the bottom-up path mesoscale. The matched feature maps are combined by horizontal and horizontal connections. In the feature pyramid network, those layers whose output feature map size can be said to be at the same stage of the network, and each stage can output an independent feature map for monitoring. In F-Dense Net, the output feature map size of each densely connected block is the same, which means that all layers in a densely connected block can be regarded as the same stage in the feature pyramid. Therefore, we choose the output of the last layer of the second, third, and fourth densely connected blocks as the reference for the feature map. This is considering that the last layer of each densely connected block has the learned features of all layers in the block, and at the same time has the richest semantic information.

In the network designed in this paper, the nearest neighbor method is used for the up-sampling of the feature map, and the up-sampling multiple is 2. In addition, we have made some improvements to the FPN method. The traditional feature pyramid horizontal connection needs to adjust the number of channels of the two feature maps to the same, and then add pixels on each channel to generate the final feature map. In this paper, combining the characteristics of the densely connected network, different horizontal connection methods are adopted. The two feature maps are merged into a new one with the same size and the same dimension as the sum of the dimensions of the two original feature maps using the feature map cascade method in the dense connection. The feature map better retains the information of the two feature maps. Finally, a 1×1 convolutional layer is added after the new feature map. On the one hand, it is to alleviate the aliasing effect caused by upsampling. It can strengthen the fusion of different channel features, while compressing the number of parameters and reducing the amount of network calculations.

Spatial Pyramid Pooling

Through research, we found that the multi-scale feature fusion in MF-Dense Net mainly focuses on the fusion between the global features of different levels of convolutional layers, and there is no effective fusion and extraction of multi-scale local features on the same convolutional layer. In response to this problem, we propose an improved spatial pyramid pooling structure to improve this problem.

The spatial pyramid pooling structure can generate fixed-size outputs from all sizes of inputs. Spatial pyramid pooling is generally connected between the last convolutional layer of the convolutional neural network and the fully connected layer to meet the requirements of the fully connected layer for the dimensionality of the input data. The convolutional layer produces an output of w*h*d, where w*h is the size of the feature map, and d is the dimension, which is determined by the number of filters in the convolutional layer. After inputting this feature map into the spatial pyramid pooling structure, after pooling with different core sizes and step lengths, three feature maps of 4*4*d, 2*2*d, and 1*1*d are obtained, and then the feature map is converted into a 1-dimensional vector and stitched together, then no matter what the size of the input feature map is, the output of the same size will be generated after the spatial pyramid pooling, that is, $1 \times 1 \times 21$d.

The spatial pooling pyramid merges the features pooled by different pooling core sizes during the operation process, and realizes the fusion of multi-scale local area features on the same layer of feature maps. However, the traditional spatial pyramid converts multiple feature map information into one-dimensional vectors and outputs them to the fully connected layer, which is not applicable in our method. Therefore, a spatial pyramid method of outputting three-dimensional feature maps is used in this article.

In the improved spatial pyramid pooling of feature maps, the input spatial pyramid pooling feature maps are pooled with multiple different pooling core sizes. In ordinary spatial pyramid pooling, the pooling step size is generally greater than 1, which plays the role of downsampling the feature map; and in this improved method, the step size of the three pooling layers used is all 1. The pooling core size can be set to 1, 1/2, and 1/3 times the edge length of the feature map (rounded up for non-integer cases). The pooling form uses maximum value pooling, and after pooling, a new feature map with the same size is the input feature map ($w \times h \times d$), and then cascades these three feature maps together in the dimension direction to obtain a $w \times h \times 3d$ feature map, which is obtained by cascading fusion of multi-scale local area features.

In terms of specific implementation, we added a spatial pooling pyramid module after the last densely connected block

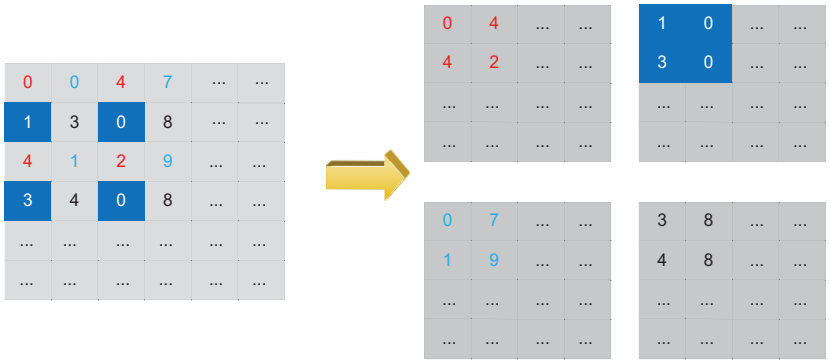


FIGURE 6 | Schematic diagram of reconstruction operation.

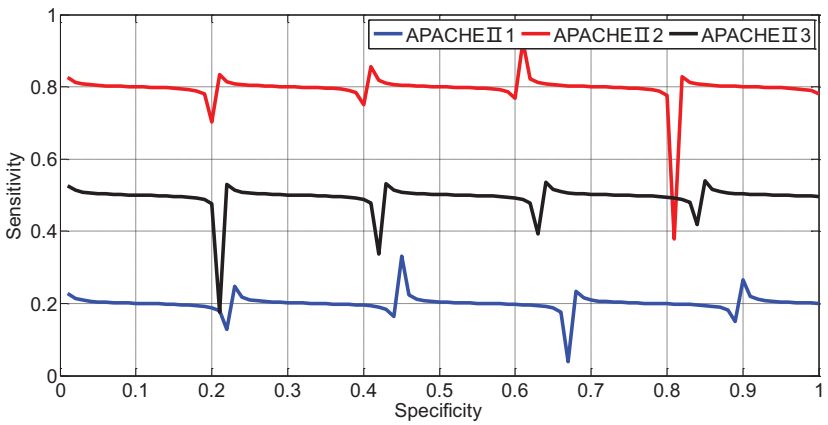


FIGURE 7 | ROC curve of APACHEII1, APACHEII2, and APACHEII3 predicting the risk of death.

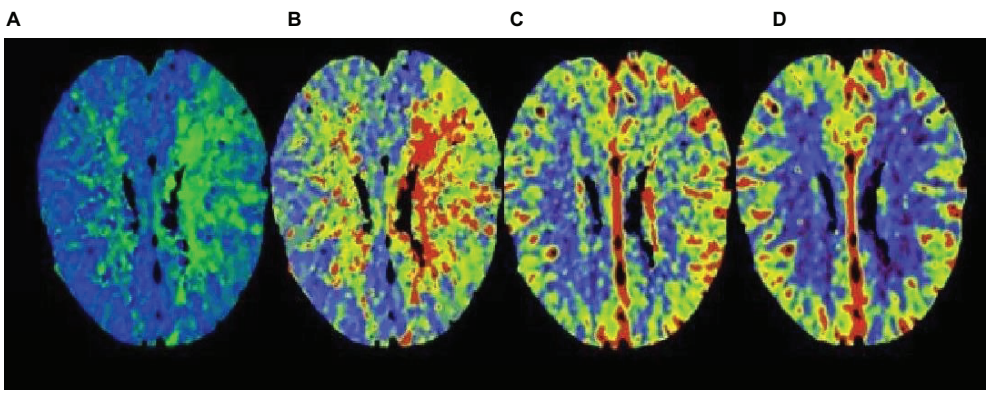


FIGURE 8 | Image performance of severe cerebrovascular disease based on multi-scale dynamic brain imaging. **(A)** Imaging manifestation of severe cerebrovascular disease in 18 h brain imaging; **(B)** Imaging manifestation of severe cerebrovascular disease in 36 h brain imaging; **(C)** Imaging manifestation of severe cerebrovascular disease in 54 h brain imaging; **(D)** Imaging manifestation of severe cerebrovascular disease in 72 h brain imaging.

of MF-Desne Net. We first use 1 × 1 convolution to reduce the dimensionality of the feature map output by the last layer of MF-Desne Net, and then input it into the improved spatial pyramid pooling structure to obtain a new feature map that incorporates multi-scale local features. Subsequently, the real parameter compression and multi-channel feature fusion of the new feature map are performed by 1 × 1 convolution. Finally, the feature map is cascaded with the original feature map output by MF-Dense Net, and the 1 × 1 convolution is used to reduce the dimension to 512 dimensions and input into the monitoring

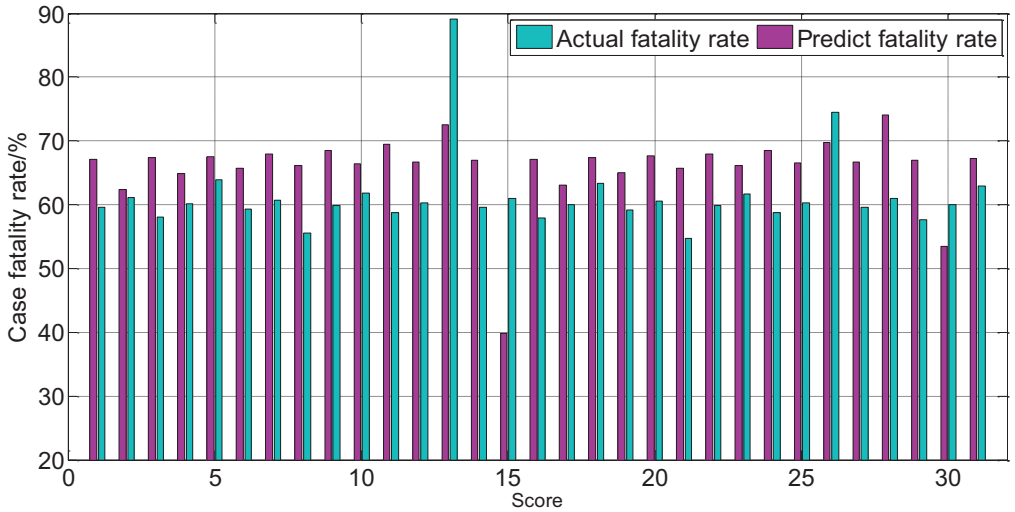


FIGURE 9 | APAHCEII1 score predicts overall mortality and actual mortality.

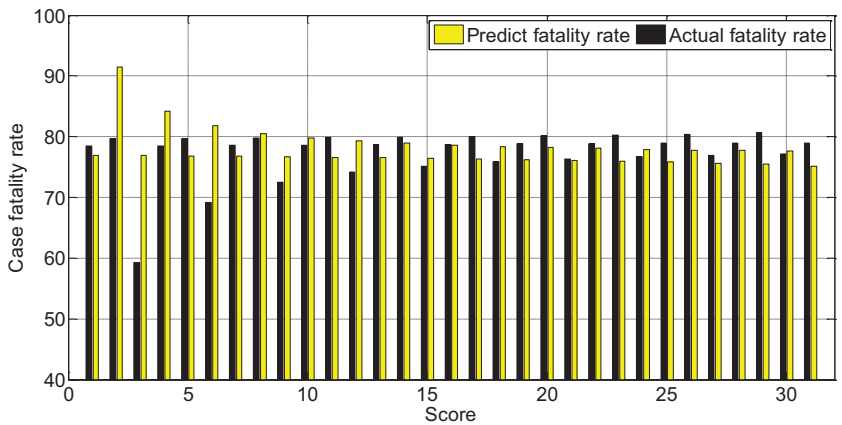


FIGURE 10 | APAHCEII2 score predicts overall mortality and actual mortality.

sub-network. By adding a spatial pooling pyramid structure, the improved backbone network can not only achieve multi-level feature fusion, but also multi-scale local feature fusion in the same dimension, and improve the accuracy of object monitoring through richer feature information.

Enhanced Utilization of Low-Level Features

Through the observation of the structure of MF-Dense Net, we found that the feature map output by the first densely connected block is not fully utilized in the feature extraction process. Because dynamic brain images often contain a large number of small-scale severe cerebrovascular diseases, the information of these severe cerebrovascular diseases is likely to be lost after multi-layer convolution operations, which makes it difficult to accurately monitor; compared to high-level feature maps, the feature map output by a densely connected block contains more low-level features and detailed information, which retains more feature patterns that can reflect small-scale severe cerebrovascular diseases and location information of some severe cerebrovascular diseases. Therefore, strengthening the use of low-level features can improve the monitoring effect of small-scale severe cerebrovascular diseases, and at the same time improve the positioning accuracy of some severe cerebrovascular diseases to be monitored. In order to make full use of the information of the feature map, we tried from different directions. The method of adding a new feature map to the top-down path increases the amount of calculation greatly and has a great impact on the monitoring speed; while the third feature map continues to be up-sampled. The method of feature fusion will also bring more extra calculations, and the aliasing effect caused by continuous up-sampling of the feature map will affect the effectiveness of the information on the high-level feature map to a certain extent. Therefore, we choose to merge the feature maps output by the first densely connected

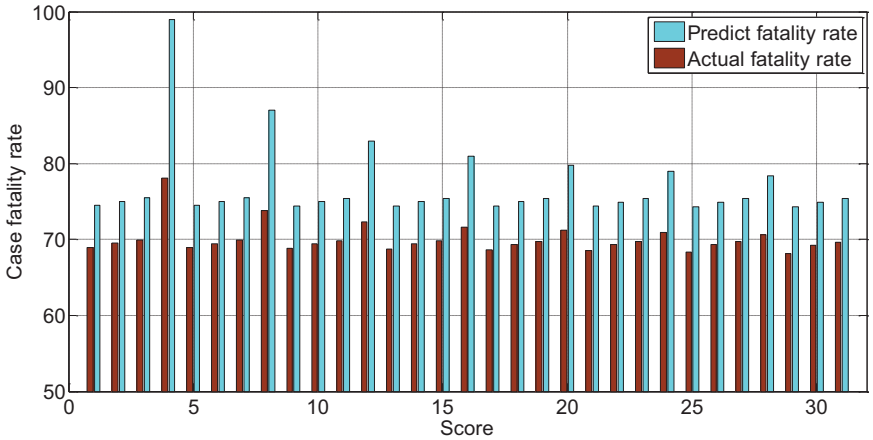


FIGURE 11 | The relationship between APAHCEII3 score predicting overall mortality rate and actual mortality rate.

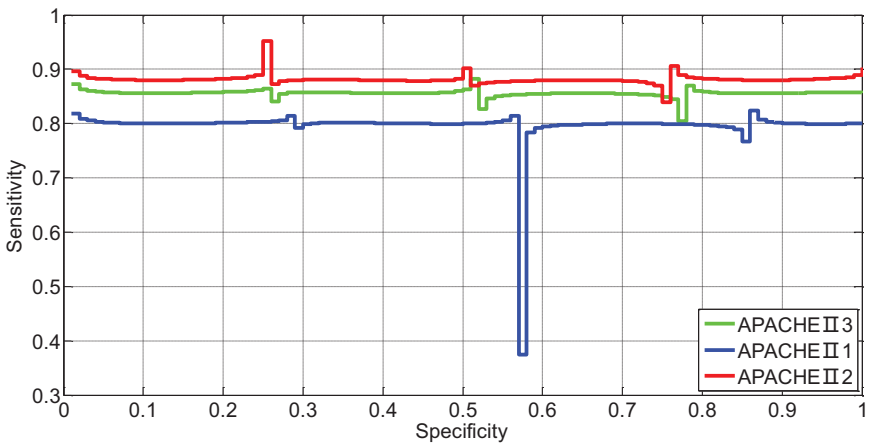


FIGURE 12 | The ROC curve of APACHEII1, APACHEII2, APACHEII3 predicting the risk of death from severe cerebral hemorrhage.

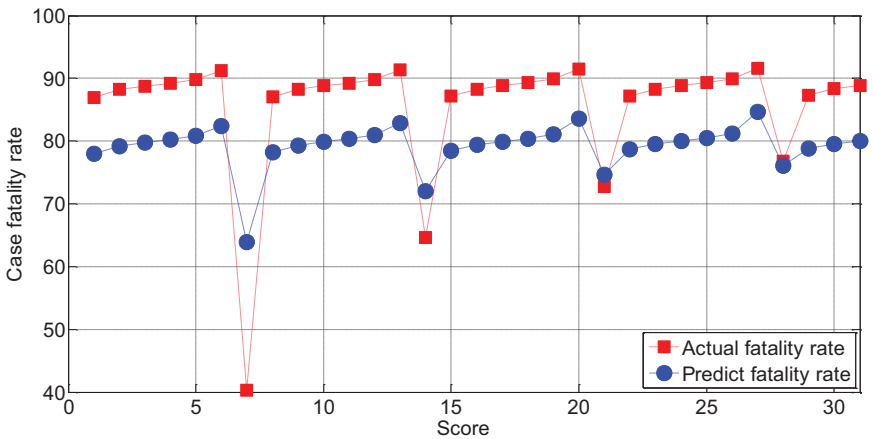


FIGURE 13 | The relationship between APACHEII1 and the mortality of patients with severe cerebral hemorrhage.

block through horizontal connection without up-sampling the third feature map.

Due to the different sizes of the two feature maps, the area of the feature map output by the first densely connected block is four times that of the third feature map in the top-down path, and the horizontal connection needs to be on several feature maps of the same size. Therefore, it is necessary to convert the larger feature map to the size of the smaller feature map to merge the two feature maps. Generally speaking, downsampling methods such as pooling can be used to reduce the size of the feature map, but this will cause a lot of loss of low-level detail information. In order to better preserve the low-level information of the feature map output by the first densely connected block, we use the Reorg Layer to reconstruct the feature map to change the feature map size, and then perform the fusion of the feature map through horizontal connections.

The reconstruction operation can change the size of the feature map without losing data. The reconstruction process is shown in **Figure 6**. Assuming that the reconstruction layer obtains a feature map input of 2w*2h*d, if the step size of the reconstruction operation is set to 2, then the reconstruction layer can convert the input feature map into a feature map output of size w*h*4d. Compared with using the pooling layer to down-sample the feature map, changing the size of the feature map through the reconstruction layer retains more low-level features, which is more conducive to the monitoring of small-scale severe cerebrovascular diseases. At the same time, these low-level features also contain more location information.

ANALYSIS OF THE RESULTS OF MONITORING AND PROGNOSIS OF SEVERE CEREBROVASCULAR DISEASE

The Relationship Between APACHEII Score and Overall Severe Cerebrovascular Disease

APACHEII1, APACHEII2, and APACHEII3 in the death group were higher than those in the survival group, and the difference between the two was statistically significant ($p < 0.01$). The change rate of APACHEII in the survival group was higher than that in the death group, and the difference between the two was statistically significant ($p < 0.01$). The area under the curve (AUR) of the ROC (Receiver Operating Curve) of APACHEII1 is 0.22; the AUR of the ROC curve of APACHEII2 is 0.83; the AUR of the ROC curve of APACHEII3 is 0.54, as shown in **Figure 7**. The accuracy of APACHEII1, APACHEII2, and APACHEII3 on the overall prognosis of patients is better, and APACHEII2 has the best accuracy.

The APACHEII1, APACHEII2, and APACHEII3 of the overall patients were divided into cut-off values of 19, 19, and 17, respectively. The score $\geq$ the cut-off value is the high-risk group, and the score $<$ cut-off value is the low-risk group. The fatality rate of the high-risk group is significantly higher than that in the low-risk group, the difference between the two groups was statistically significant.

The APACHE II that showed an upward trend within 72 h was divided into high-risk group, and those with a downward trend were divided into low-risk group. The fatality rate of the high-risk group was 67.86%, the fatality rate of the low-risk group was 17.70%, and the fatality rate of the high-risk group was significantly higher than that of the low-risk group. The difference between the two groups was statistically significant. The imaging manifestations of severe cerebrovascular diseases based on multi-scale dynamic brain imaging are shown in **Figure 8**.

The APACHEII1 score is correlated with the patient's prognosis ($r = -0.47$, $p < 0.01$). The predictive value of the APACHEII1 score for the overall mortality of severe cerebrovascular disease is generally higher than the actual mortality, as shown in **Figure 9**. APACHEII2 is correlated with prognosis ($r = -0.63$, $p < 0.01$), as shown in **Figure 10**. APACHEII3 is correlated with prognosis ($r = -0.54$, $p < 0.01$). The higher the score, the higher the mortality rate. The predicted value of APACHEII3 for the overall mortality rate is higher than the actual mortality rate, as shown in **Figure 11**.

The Relationship Between APACHEII Score and Severe Cerebral Hemorrhage

The APACHEII1, APACHEII2, and APACHEII3 scores of patients in the death group were higher than those in the survival group, and the difference between the two was statistically significant ($p < 0.01$). The change rate of APACHEII in the survival group was higher than that in the death group, and the difference between the two was statistically significant ($p < 0.01$).

The AUR of the ROC curve of APACHEII1 is 0.81; the AUR of the ROC curve of APACHEII2 is 0.88; the AUR of the ROC curve of APACHEII3 is 0.85. This shows that APACHEII1, APACHEII2, and APACHEII3 have good prognostic accuracy, and APACHEII2 has the best accuracy, as shown in **Figure 12**.

The scores corresponding to the Youden index of APACHEII1, APACHEII2, and APACHEII3 are 18, 19, and 18 points, respectively, indicating that for patients with severe cerebral hemorrhage, the first 24 h after admission to the NICU are higher than 18 points, and the internal score was higher than 19 points, and the score did not drop below 18 points in the third 24 h, and the mortality rate was high.

The APACHEII1, APACHEII2, and APACHEII3 of the overall patients were divided into cut-off values of 18, 19, and 18, respectively. The score $\geq$ the cut-off value is the high-risk group, and the score $<$ cut-off value is the low-risk group, and the fatality rate of the high-risk group is significantly higher. In the low-risk group, the difference between the two groups was statistically significant. The APACHE II that showed an upward trend within 72 h was divided into a high-risk group, and those with a downward trend were divided into a low-risk group. The fatality rate of the high-risk group was significantly higher than that of the low-risk group, and the difference between the two groups was statistically significant.

The score of APACHEII1 is correlated with the prognosis of patients ($r = -0.65$, $p < 0.01$). The higher the score, the higher the risk of death. Overall, the predictive value of the APACHEII1

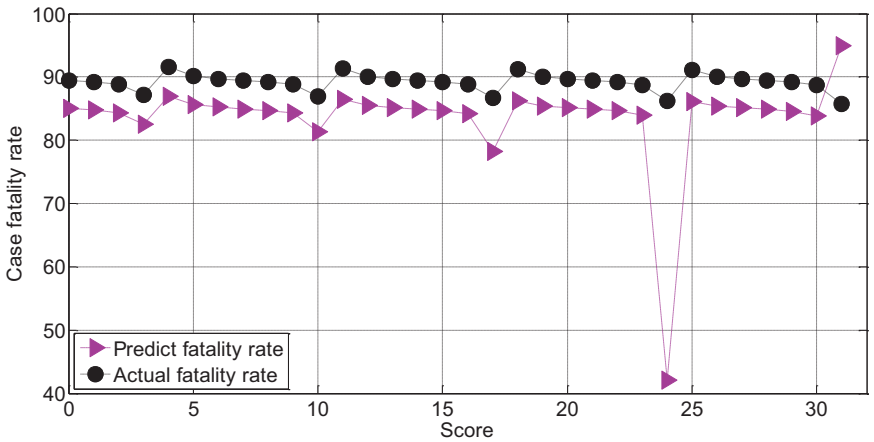


FIGURE 14 | APAHCEII2 predicts the relationship between mortality and actual mortality in patients with severe cerebral hemorrhage.

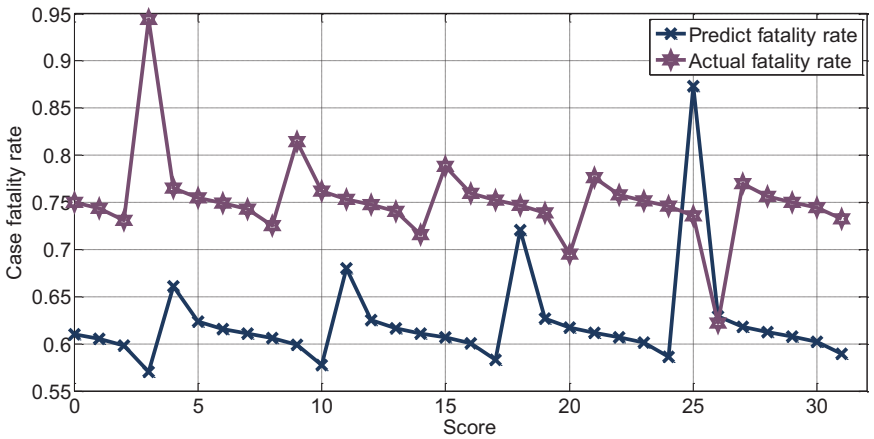


FIGURE 15 | APAHCEII3 predicts the relationship between mortality and actual mortality in patients with cerebral hemorrhage.

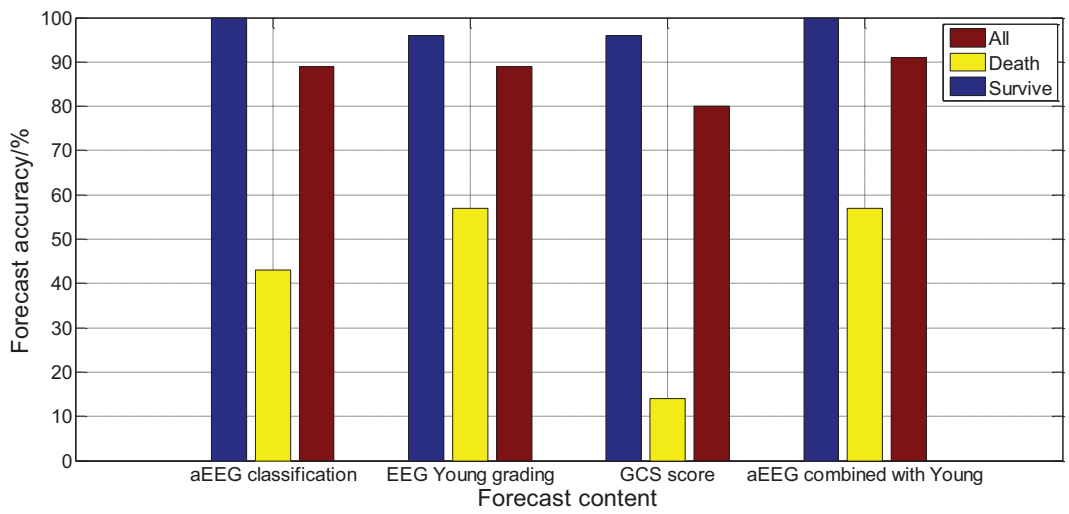


FIGURE 16 | Comparison of evaluation methods and prognostic logistic regression analysis.

score for the fatality rate of severe cerebral hemorrhage is lower than the actual value of the fatality rate, as shown in **Figure 13**.

The APACHEII2 score is correlated with the patient's prognosis (r = -0.78, $p < 0.01$). The higher the APACHEII2 score, the higher the risk of death. The predicted value of the fatality rate for severe cerebral hemorrhage with APACHEII2 score below 30 points is always lower than the actual fatality rate. The actual mortality rate is lower than the predicted mortality rate above 30 points, as shown in **Figure 14**.

The score of APACHEII3 was correlated with the prognosis of patients ($r = -0.72$, $p < 0.01$). The APACHEII3 score is less than 24 points, and the predicted value of the fatality rate of severe cerebral hemorrhage is always lower than the actual fatality rate. The actual mortality rate is lower than the predicted mortality rate at 24–26 points, as shown in **Figure 15**.

Logistic Regression Analysis and Comparison of Evaluation Methods and Prognosis

Logistic regression analysis was performed on the aEEG (Ambulatory Electroencephalogram) classification and prognosis, and the χ^2 test was performed on the fit of the regression equation model, $P = 0.004$, indicating that the model was well fit and the regression equation was significant. Logistic regression analysis showed that a EEG classification had an accuracy rate of 100% for survival prognosis, an accuracy rate of 42.5% for death prognosis, and an accuracy rate of 87.1% for comprehensive prognosis. Logistic regression analysis was performed on the Young grading and prognosis of EEG, and the χ^2 test was performed on the fit of the regression equation model, $P = 0.003$, indicating that the fit of the model was good and the regression equation was significant. Logistic regression analysis showed that the accuracy of EEG Young classification for survival prognosis judgment was 97.7%, the accuracy of death prognosis judgment was 58.9%, and the accuracy of comprehensive prognosis judgment was 89.1%. Logistic regression analysis was performed on the GCS score and prognosis, and the fit of the regression equation model was tested by χ^2, $P = 0.016$, indicating that the model fit is acceptable and the regression equation is meaningful. Logistic regression analysis showed that the accuracy of GCS score for survival prognosis judgment was 97.2%, the accuracy of death prognosis judgment was 13.9%, and the accuracy of comprehensive prognosis judgment was 79%, as shown in **Figure 16**. The results show that aEEG classification and EEG Young classification are more accurate in prognostic judgment than GCS score; and aEEG classification combined with EEG Young classification can improve the prediction accuracy.

CONCLUSION

Severe cerebrovascular disease has an extremely high disability rate and mortality rate, and correct assessment of its prognosis is of great help to adopt correct treatments and reduce the disability rate and mortality rate. In today's era of highly developed imaging, the majority of patients with cerebrovascular disease are the first choice for cranial computed tomography (CT) or magnetic resonance imaging (MRI) to assess the patient's prognosis through objective images. This article studies the correlation between long-term EEG classification and typical EEG configuration and patient prognosis, in order to find a more objective clinical method for evaluating the condition and prognosis of patients with severe cerebrovascular disease. In this paper, a study on the feature extraction of critical cerebrovascular disease monitoring tasks based on multi-scale dynamic brain imaging is carried out. In view of the large-scale changes and diverse characteristics of severe cerebrovascular diseases in dynamic brain images, the idea of improving the monitoring effect through multi-scale fusion of the characteristics of severe cerebrovascular diseases is proposed. We designed a fast and robust dense connection network F-Desne Net with excellent feature extraction capabilities, and combined with the feature gold tower network architecture, we built a severe cerebrovascular disease that can extract multi-layer fusion features and output multi-scale feature maps. Based on MF-Desne Net, a fast and accurate end-to-end dynamic brain imaging monitoring algorithm for severe cerebrovascular disease MDRD is designed. In addition, we designed an improved method on the basis of MDRD. Improved feature pyramid pooling is added to improve the network's multi-scale local feature fusion ability; the use of low-level features of the network is enhanced to improve the network's ability to extract basic feature patterns and detailed information of the image. The improved method can effectively improve the monitoring performance of the MDRD method. This study found that APACHEII1, APACHEII2, and APACHEII3 in the death group were higher than those in the survival group. The change rate of APACHEII in the survival group was higher than that in the death group ($p < 0.01$). APACHEII in different periods was correlated with prognosis. APACHEII has a good predictive value for the prognosis of severe cerebrovascular disease, and has clinical practice value for predicting the risk of death. Dynamic observation of the changes in APACHEII score is of significance for the prediction and prognostic evaluation of severe cerebrovascular disease in neurology.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://brainweb.bic.mni.mcgill.ca/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Hanyang Hospital, Wuhan University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

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

REFERENCES

Conzen, C., Becker, K., Albanna, W., Weiss, M., Bach, A., Lushina, N., et al. (2019). The acute phase of experimental subarachnoid hemorrhage: intracranial pressure dynamics and their effect on cerebral blood flow and autoregulation. *Transl. Stroke Res.* 10, 566–582. doi: 10.1007/s12975-018-0674-3

Dehkharghani, S., and Qiu, D. (2020). MR thermometry in cerebrovascular disease: physiologic basis, hemodynamic dependence, and a new frontier in stroke imaging. *Am. J. Neuroradiol.* 41, 555–565. doi: 10.3174/ajnr.a6455

Dumitrascu, O. M., and Koronyo-Hamaoui, M. (2020). Retinal vessel changes in cerebrovascular disease. *Curr. Opin. Neurol.* 33, 87–92. doi: 10.1097/wco.0000000000000779

Forti, R. M., Favilla, C. G., Cochran, J. M., Baker, W. B., Detre, J. A., Kasner, S. E., et al. (2019). Transcranial optical monitoring of cerebral hemodynamics in acute stroke patients during mechanical thrombectomy. *J. Stroke Cerebrovasc. Dis.* 28, 1483–1494. doi: 10.1016/j.jstrokecerebrovasdis.2019.03.019

Khalil, A. A., Villringer, K., Filleböck, V., Hu, J. Y., Rocco, A., Fiebach, J. B., et al. (2020). Non-invasive monitoring of longitudinal changes in cerebral hemodynamics in acute ischemic stroke using BOLD signal delay. *J. Cerebr. Blood Flow Metab.* 40, 23–34. doi: 10.1177/0271678x18803951

Koskinen, L. O. D., Sundström, N., Hägglund, L., Eklund, A., and Olivecrona, M. (2019). Prostacyclin affects the relation between brain interstitial glycerol and cerebrovascular pressure reactivity in severe traumatic brain injury. *Neurocritical Care* 31, 494–500. doi: 10.1007/s12028-019-00741-4

Li, B., Zhao, M., Feng, L., Dou, C., Ding, S., Zhou, G., et al. (2020). Organic NIR-II molecule with long blood half-life for in vivo dynamic vascular imaging. *Nat. Commun.* 11, 1–11.

Liu, C. H., Abrams, N. D., Carrick, D. M., Chander, P., Dwyer, J., Hamlet, M. R. J., et al. (2019). Imaging inflammation and its resolution in health and disease: current status, clinical needs, challenges, and opportunities. *FASEB J.* 33, 13085–13097. doi: 10.1096/fj.201902024

Lohith, T. G., Bennacef, I., Vandenberghe, R., Vandenbulcke, M., Salinas, C. A., Declercq, R., et al. (2019). Brain imaging of Alzheimer dementia patients and elderly controls with 18F-MK-6240, a PET tracer targeting neurofibrillary tangles. *J. Nuclear Med.* 60, 107–114. doi: 10.2967/jnumed.118.208215

Maïer, B., Dargazanli, C., Bourcier, R., Kyheng, M., Labreuche, J., Mosimann, P. J., et al. (2020). Effect of steady and dynamic blood pressure parameters during thrombectomy according to the collateral status. *Stroke* 51, 1199–1206. doi: 10.1161/strokeaha.119.026769

Malas, M. B., Lorenzo, J. I. L., Nejim, B., Hanover, T. M., Mehta, M., Kashyap, V., et al. (2019). Analysis of the ROADSTER pivotal and extended-access cohorts shows excellent 1-year durability of transcarotid stenting with dynamic flow reversal. *J. Vasc. Surg.* 69, 1786–1796. doi: 10.1016/j.jvs.2018.08.179

Marini, D., Xu, J., Sun, L., Jaeggi, E., and Seed, M. (2020). Current and future role of fetal cardiovascular MRI in the setting of fetal cardiac interventions. *Prenat. Diagn.* 40, 71–83. doi: 10.1002/pd.5626

Mathieu, F., Zeiler, F. A., Whitehouse, D. P., Das, T., Ercole, A., Smielewski, P., et al. (2020). Relationship between measures of cerebrovascular reactivity and intracranial lesion progression in acute TBI patients: an exploratory analysis. *Neurocrit. Care* 32, 373–382. doi: 10.1007/s12028-019-00885-3

Merikangas, K. R., Swendsen, J., Hickie, I. B., Cui, L., Shou, H., Merikangas, A. K., et al. (2019). Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. *JAMA Psychiatry* 76, 190–198. doi: 10.1001/jamapsychiatry.2018.3546

Quon, J. L., Kim, L. H., Lober, R. M., Maleki, M., Steinberg, G. K., and Yeom, K. W. (2019). Arterial spin-labeling cerebral perfusion changes after revascularization surgery in pediatric moyamoya disease and syndrome. *J. Neurosurg. Pediatr.* 23, 486–492. doi: 10.3171/2018.11.peds18498

Rahmim, A., Lodge, M. A., Karakatsanis, N. A., Panin, V. Y., Zhou, Y., McMillan, A., et al. (2019). Dynamic whole-body PET imaging: principles, potentials and applications. *Eur. J. Nuclear Med. Mol. Imaging* 46, 501–518. doi: 10.1007/s00259-018-4153-6

Robertson, A. D., Atwi, S., Kostoglou, K., Verhoeff, N. P. L. G., Oh, P. I., Mitsis, G. D., et al. (2019). Cerebrovascular pulsatility during rest and exercise reflects hemodynamic impairment in stroke and cerebral small vessel disease. *Ultrasound Med. Biol.* 45, 3116–3127. doi: 10.1016/j.ultrasmedbio.2019.08.019

Rocha, E. A., and Singhal, A. B. (2020). Current treatment options in cardiovascular medicine: update on reversible cerebral vasoconstriction syndrome. *Curr. Treat. Options Cardiovasc. Med.* 22, 1–13.

Rosenberg, A. J., Schroeder, E. C., Grigoriadis, G., Wee, S. O., Bunsawat, K., Heffernan, K. S., et al. (2020). Aging reduces cerebral blood flow regulation following an acute hypertensive stimulus. *J. Appl. Physiol.* 128, 1186–1195. doi: 10.1152/japplphysiol.00137.2019

Savić, T., Gambino, G., Bokharaie, V. S., Noori, H. R., Logothetis, N. K., and Angelovski, G. (2019). Early detection and monitoring of cerebral ischemia using calcium-responsive MRI probes. *Proc. Natl. Acad. Sci. U.S.A.* 116, 20666–20671. doi: 10.1073/pnas.1908503116

Tanaka, T., Stephenson, M. C., Nai, Y. H., Khor, D., Saridin, F. N., Hilal, S., et al. (2020). Improved quantification of amyloid burden and associated biomarker cut-off points: results from the first amyloid Singaporean cohort with overlapping cerebrovascular disease. *Eur. J. Nuclear Med. Mol. Imaging* 47, 319–331. doi: 10.1007/s00259-019-04642-8

Taneja, K., Lu, H., Welch, B. G., Thomas, B. P., Pinho, M., Lin, D., et al. (2019). Evaluation of cerebrovascular reserve in patients with cerebrovascular diseases using resting-state MRI: a feasibility study. *Magn. Resonan. Imaging* 59, 46–52. doi: 10.1016/j.mri.2019.03.003

Václavů, L., Meynart, B. N., Mutsaerts, H. J. M. M., Petersen, E. T., Majoie, C. B. L. M., VanBavel, E. T., et al. (2019). Hemodynamic provocation with acetazolamide shows impaired cerebrovascular reserve in adults with sickle cell disease. *Haematologica* 104, 690–699. doi: 10.3324/haematol.2018.206094

Waddle, S. L., Juttukonda, M. R., Lants, S. K., Davis, L. T., Chitale, R., Fusco, M. R., et al. (2020). Classifying intracranial stenosis disease severity from functional MRI data using machine learning. *J. Cerebr. Blood Flow Metab.* 40, 705–719. doi: 10.1177/0271678x19848098

Wang, C., Keliher, E., Zeller, M. W. G., Wojtkiewicz, G. R., Aguirre, A. D., Buckbinder, L., et al. (2019). An activatable PET imaging radioprobe is a dynamic reporter of myeloperoxidase activity in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 116, 11966–11971.

Wang, Y., Nencka, A. S., Meier, T. B., Guskiewicz, K., Mihalik, J. P., Alison Brooks, M., et al. (2019). Cerebral blood flow in acute concussion: preliminary ASL findings from the NCAA-DoD CARE consortium. *Brain Imaging Behav.* 13, 1375–1385. doi: 10.1007/s11682-018-9946-5

Wardlaw, J. M., Smith, C., and Dichgans, M. (2019). Small vessel disease: mechanisms and clinical implications. *Lancet Neurol.* 18, 684–696. doi: 10.1016/s1474-4422(19)30079-1

Xu, X., Hua, Y., Wang, L., Hou, W., and Xia, M. (2020). Correlation between risk factors of cerebrovascular disease and calcified plaque characteristics in patients with atherosclerotic severe carotid stenosis. *Neurol. Res.* 42, 83–89. doi: 10.1080/01616412.2019.1710403

Zafar, A., Quadri, S. A., Farooqu, I. M., Ikram, A., Robinson, M., Hart, B. L., et al. (2019). Familial cerebral cavernous malformations. *Stroke* 50, 1294–1301.

Zeiler, F. A., Ercole, A., Czosnyka, M., Smielewski, P., Hawryluk, G., Hutchinson, P. J. A., et al. (2020). Continuous cerebrovascular reactivity monitoring in moderate/severe traumatic brain injury: a narrative review of advances in neurocritical care. *Br. J. Anaesthesia* 124, 440–453. doi: 10.1016/j.bja.2019.11.031

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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 July 2021
doi: 10.3389/fnins.2021.705323

Semi-Supervised Support Vector Machine for Digital Twins Based Brain Image Fusion

*Zhibo Wan[1], Youqiang Dong[2], Zengchen Yu[1], Haibin Lv[3] and Zhihan Lv[1]**

[1] College of Computer Science and Technology, Qingdao University, Qingdao, China, [2] R&D Department, Qingdao Haily Measuring Technologies Co., Ltd., Qingdao, China, [3] North China Sea Offshore Engineering Survey Institute, Ministry Of Natural Resources North Sea Bureau, Qingdao, China

The purpose is to explore the feature recognition, diagnosis, and forecasting performances of Semi-Supervised Support Vector Machines (S3VMs) for brain image fusion Digital Twins (DTs). Both unlabeled and labeled data are used regarding many unlabeled data in brain images, and semi supervised support vector machine (SVM) is proposed. Meantime, the AlexNet model is improved, and the brain images in real space are mapped to virtual space by using digital twins. Moreover, a diagnosis and prediction model of brain image fusion digital twins based on semi supervised SVM and improved AlexNet is constructed. Magnetic Resonance Imaging (MRI) data from the Brain Tumor Department of a Hospital are collected to test the performance of the constructed model through simulation experiments. Some state-of-art models are included for performance comparison: Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), AlexNet, and Multi-Layer Perceptron (MLP). Results demonstrate that the proposed model can provide a feature recognition and extraction accuracy of 92.52%, at least an improvement of 2.76% compared to other models. Its training lasts for about 100 s, and the test takes about 0.68 s. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of the proposed model are 4.91 and 5.59%, respectively. Regarding the assessment indicators of brain image segmentation and fusion, the proposed model can provide a 79.55% Jaccard coefficient, a 90.43% Positive Predictive Value (PPV), a 73.09% Sensitivity, and a 75.58% Dice Similarity Coefficient (DSC), remarkably better than other models. Acceleration efficiency analysis suggests that the improved AlexNet model is suitable for processing massive brain image data with a higher speedup indicator. To sum up, the constructed model can provide high accuracy, good acceleration efficiency, and excellent segmentation and recognition performances while ensuring low errors, which can provide an experimental basis for brain image feature recognition and digital diagnosis.

Keywords: semi-supervised support vector machines, brain image, digital twins, image segmentation, improved AlexNet

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing Medical University, China
Lu Wang,
Shandong Agricultural University, China

****Correspondence:***
Zhihan Lv
lvzhihan@gmail.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *05 May 2021*
Accepted: *07 June 2021*
Published: *09 July 2021*

Citation:
Wan Z, Dong Y, Yu Z, Lv H and Lv Z (2021) Semi-Supervised Support Vector Machine for Digital Twins Based Brain Image Fusion. Front. Neurosci. 15:705323. doi: 10.3389/fnins.2021.705323

INTRODUCTION

Now that science and technology advances quickly, Artificial Intelligence (AI), Big Data (BD), the Internet of Things (IoT), and Deep Learning (DL) have been accepted in all walks of life. Health is vital to the happy life of people. Medical intelligentization has improved health monitoring techniques notably. Brain tumors have severely harmed people's health.

Statistics suggest that the annual mortality rate of brain cancer in China is second only to gastric cancer and lung cancer. About 80,000 people die from brain cancer each year, accounting for about 22.5% of the world's total number of brain cancers (Zhou et al., 2019; Kang et al., 2021). In only a few decades of the 21st century, the number of people diagnosed with brain cancer continues to increase, especially in Asian countries (Zheng et al., 2020). Thus, brain health has become the focus of brain disease experts, researchers, and brain surgeons.

Medical imaging techniques play a vital role in clinical diagnosis with the successive introduction of medical imaging equipment and the continuous development of computer science and technology. Results of brain disease diagnosis depend on brain image acquisition and accurate interpretation. However, due to the large number of patient groups, the number of brain and other medical images keeps increasing dramatically (Tobon Vasquez et al., 2020). Manual image reading is apparently inefficient, and the workload for clinicians is unbearable. Therefore, using computer-aided diagnosis to extract features from medical images is of great significance in improving diagnostic accuracy and reducing clinicians' workload.

Image segmentation is one of the necessary steps of brain image fusion, so it is necessary to segment and extract the features of the image. There are many approaches to extract brain image features; currently trendy ones include Region Growing (RG), Level Set, Fuzzy Clustering, and Machine Learning (ML) (Aheleroff et al., 2020; Angin et al., 2020; Li et al., 2020; Pan et al., 2021). However, these approaches require manual image segmentation, consuming lots of human resources and materials. As a burgeoning AI algorithm, ML extracts data features and investigates data relationships through self-learning on brain images, which dramatically increases image segmentation accuracy and stability while improving efficiency. As Digital Twins (DTs) evolve, their applications in brain image segmentation have enhanced data generation speed and scale, leading to a remarkable increase in big data processing complexity and brain image analysis (Gudigar et al., 2019). Digital twins can map complex brain images from real space to virtual space, and it is of great significance to extract features, segment, and fuse images.

To sum up, AI applications in medical fields, such as brain image processing, can help clinicians choose surgical treatment or radiotherapy and chemotherapy treatment options for patients while avoiding the steps of direct surgical resection of all tumors, which greatly reduces the risk of treatment. The innovative points are (1) the Semi-Supervised Support Vector Machines (S3VMs) proposed regarding the massive volume of unlabeled brain image data, (2) the improved AlexNet, and (3) the brain image fusion DTs diagnosis and forecasting model constructed based on S3VMs and improved AlexNet, in an effort to provide theoretical support for brain image feature recognition and digital diagnosis.

RECENT WORKS

ML Applications

Now that complex data grow explosively, big data mining and analysis functions, especially DL, are demanded increasingly, which has been explored by many scholars. Yan et al. (2018) put forward Device ElectroCardioGram (DECG) and constructed a forecasting algorithm for the remaining service life of industrial equipment based on Deep Denoising Autoencoder (DDA) and regression calculation. Comparison with traditional factory information models validated the feasibility and effectiveness of the proposed algorithm (Yan et al., 2018). Wu et al. (2019) adopted deep learning algorithms in medical imaging to understand patients' physical condition, thereby improving the disease treatment. Ghosh et al. (2019) developed a DL approach for molecular excitation spectroscopy forecasting. They trained and evaluated Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Deep Tensor Neural Network (DTNN) to analyze the electronic state density of organic molecules. Eventually, they discovered that real-time spectroscopic forecasting could be performed on the structure of tiny organic molecules, and potential application molecules could be determined (Ghosh et al., 2019). Shao and Chou (2020) developed a CNN-based forecasting factor for the subcellular localization of viral proteins, called Ploc-Deep-mVirus, for the Coronavirus Disease 2019 (COVID-19) currently spreading globally. They found that this forecasting factor was particularly suitable for processing multi-site systems, and its forecasting performance was remarkably better than other advanced forecasting indicators (Shao and Chou, 2020). Ahmed et al. (2020) applied DL approaches to the clinical or behavioral recognition of Autism Spectrum Disorder (ASD). They designed an image generator, which could produce a single-volume brain image from the entire brain image by separately considering the voxel time point of each subject. Eventually, performances of four different DL approaches, as well as their corresponding integrated classifiers and the algorithms, were assessed. Results demonstrated a better effect on a large-scale multi-site brain imaging dataset (bedict) (Ahmed et al., 2020).

Research Progress of Image Fusion DTs

Qi and Tao (2018) applied DTs to image information physical integration of the manufacturing industry. They compared the similarities and differences between BD and DTs from the overall and data perspectives. Consequently, they discussed using DTs to promote smart manufacturing (Qi and Tao, 2018). As embedded sensors, BD processing, and cloud computing got advanced, Barricelli et al. (2019) applied DTs to medical image diagnosis and recognition, and consequently, validated its effectiveness and feasibility. Liu et al. (2019) put forward a Cloud DTs Healthcare (CloudDTH) architecture to monitor, diagnose, and forecast image data collected by wearable

medical devices to forecast individual health status, thereby achieving individual health management, especially the health management of seniors. Li et al. (2021) proposed a novel semantic-enhanced DTs system for accurate and fine-grained real-time monitoring of Robot-Environment Interaction (REI). This system adopted the robot's external perception and ontology perception as the mechanism of interaction state monitoring and semantic reasoning. Meanwhile, they proposed a multi-feature fusion visual relationship detector and introduced a lightweight spatial relationship dataset to measure the visual relationship between entity pairs. They evaluated the real-time monitoring ability and semantic reasoning ability of the system and verified the effectiveness and feasibility of the scheme (Li et al., 2021).

The above works reveal that most ML algorithms applied in the medical field collect image data from historical cases. In contrast, DTs are mappings from virtual space to real cases. There are very few combinations of DTs and ML in the medical field. Hence, ML approaches and DTs are introduced for brain image fusion in the present work, which is of great significance to the accurate identification and diagnosis of brain images.

ML-Based Brain Image Fusion DTs Diagnosis and Forecasting

At present, although many achievements have been made in the fusion of medical images such as brain tumors, the results are still unsatisfactory, and the application effect in clinical medicine is extremely limited. As one of the effective methods of brain image fusion, image segmentation technology has great practical value and research significance in the diagnosis, prediction, and feature extraction of brain tumor and other medical images.

Demand Analysis of Brain Image Fusion DTs Diagnosis and Forecasting

With the aging of the population, the number of patients is increasing, and the number of medical images is growing. It has become a new direction to diagnose and treat diseases by using computer technology such as digital twins to assist the recognition and prediction of brain images. Sun et al. (2020) at present, common medical imaging techniques include X-ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound Imaging (UI), Positron Emission Tomography (PET), and Molecular Imaging (MI) (Wang et al., 2020). MRI is very suitable for brain tumor research and diagnosis because of its non-invasiveness, high soft-tissue contrast, high resolution, and no direction limitations compared with other imaging techniques (Liu Y. et al., 2018).

Regarding brain image processing, excellent brain image segmentation results are of great significance for further target feature extraction and target recognition and analysis. Brain image segmentation differs from ordinary image segmentation in image complexity and diversity (Wang et al., 2018). Brain image segmentation is a crucial branch of medical image segmentation. Key difficulties are as follows: (1) the gray scale distribution between different tissues is not uniform, and the distribution of different tissues overlaps with each other. (2) Brain tissues have complicated structures. Generally, brain MRI images include the brain cortex, gray matter, white matter, cerebrospinal fluid, and other tissues; each has a complicated shape and structure. (3) Boundaries of brain tumors are blurry, and the density distribution is uneven. Thus, in the analysis of brain image processing, the brain image and structure in real space can be mapped to virtual space by digital twins. After the virtual brain image feature extraction and image fusion processing, the treatment scheme for patients with brain diseases can be formulated, and the treatment purpose for patients can be achieved finally. Present brain image segmentation approaches include supervised learning, unsupervised learning, and DL, as illustrated in **Figure 1**.

To improve the effect of brain image segmentation diagnosis, both unlabeled and labeled data are adopted to put forward an S3VMs regarding the massive volume of unlabeled brain image data based on the above approaches of brain image segmentation diagnosis. S3VMs combine supervised learning and unsupervised learning, which can utilize labeled data and unlabeled data simultaneously to improve the model's generalization ability (Liu L. et al., 2018; Zavrak and İskefiyeli, 2020). Combining semi-supervised learning and Support Vector Machines (SVMs) based on labeled samples obtained by time-domain simulation and the unlabeled samples from the actual system to assess brain image data can extract actual data distribution comprehensively and improve the weak generalization ability of present supervised feature extraction and assessment models, in an effort to enhance model's adaptability to the actual system. This idea can solve the shortcomings of the current supervised ML feature extraction models and the starting point of the present work.

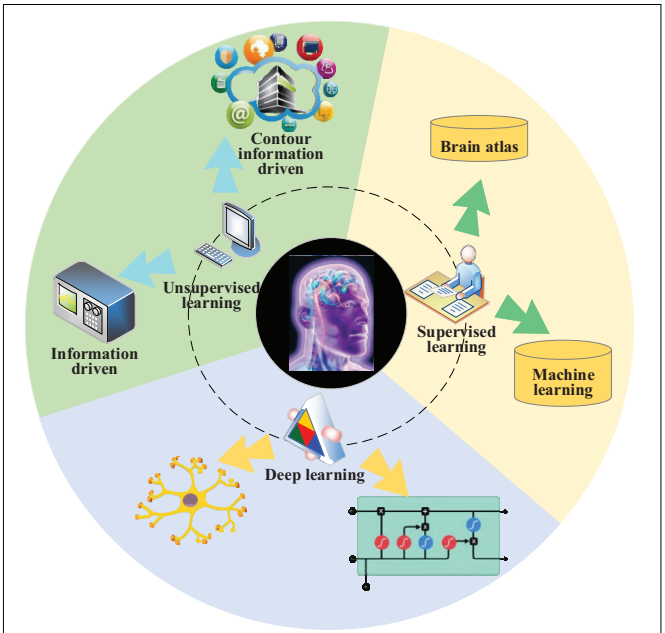


FIGURE 1 | Schematic diagram of brain image segmentation fusion diagnosis.

Performance of S3VMs in Brain Image Feature Extraction

Semi-supervised learning combines supervised learning and unsupervised learning. It utilizes both labeled data and unlabeled data for learning and mine the data distribution hidden in unlabeled samples such as brain images, thereby guiding the training of the classifier and improving its performance.

A graph-based semi-supervised learning algorithm is adopted to express the relationships between the entire training data during brain image feature extraction. Graph nodes represent training samples, including labeled samples and unlabeled samples. There are edges between nodes; these edges are given weights to represent the similarity between data nodes and can be obtained using various distance metrics. Overall, the greater the weight, the greater the similarity. No connection between nodes means zero similarity between them. Similar vertices are given the same category labels as much as possible so that the graph distribution is as smooth as possible.

The mathematical model of standard SVM can be formalized as a convex quadratic programming problem; then, a linear classification machine can be obtained through hard interval maximization learning (Zhang et al., 2020). If the training data are non-linear and inseparable, the kernel function will be utilized to map the input space to the feature space, and the linear SVM is implicitly learned in the high-dimensional feature space by maximizing the soft interval. The linearly separable optimal classification hyperplane is demonstrated in **Figure 2**.

As shown in **Figure 2**, a linearly separable sample set $T = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ is given, where (x_i, y_i) refers to the sample point, $x_i \in R^n$, and $y_i \in \{+1, -1\}, i = 1, 2, \cdots, n$. The red and the black circles represent two types of training samples, respectively. At this point, there is an infinite number of interfaces that can correctly separate the two types of samples. Hence, an optimal classification hyperplane is required, which should not only ensure the minimum empirical risk and correctly divide the two types of samples but also minimize the actual risk and meet the maximum classification interval. The core idea of linear separable SVM employs the maximum classification interval to find the optimal hyperplane, at which the solution is unique. In **Figure 2**, H denotes the optimal classification hyperplane, H1 and H2 are the planes containing the closest sample points (support vectors) to the classification hyperplane. H1 and H2 are parallel to H, and the distance between them is called the classification interval. The classification hyperplane function of linear separable SVM is:

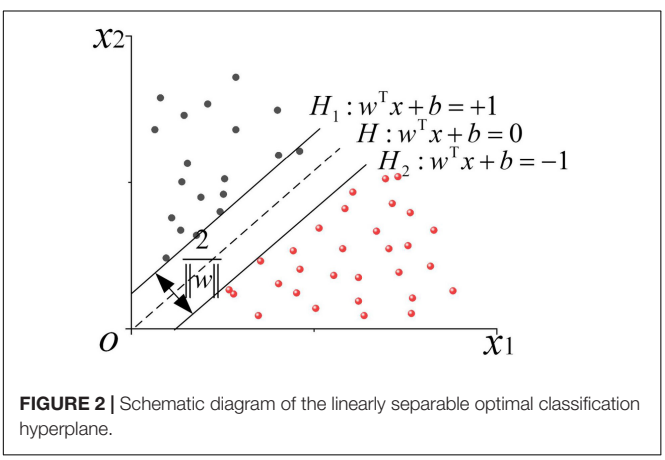


FIGURE 2 | Schematic diagram of the linearly separable optimal classification hyperplane.

$$w \cdot x + b = 0 \tag{1}$$

In Eq. 1, w represents the normal vector of the hyperplane, and b denotes the bias. The following conditions must be met to classify samples accurately and ensure the maximum interval simultaneously:

$$y_i \left(w \cdot x + b\right) \geq 1 \tag{2}$$

The classification interval can be obtained as 2/1I y} II. Solving the optimal hyperplane of the training sample set T can be transformed into a problem described by the following equation:

$$\begin{aligned} &\min_{a,b} \frac{\| w \|^2}{2} \\ &s.t.\ y_i \left(w \cdot x + b\right) \geq 1, i = 1, 2, \cdots, n \end{aligned} \tag{3}$$

Equation (3) is a quadratic programming problem, which can be solved as a dual Lagrangian problem. Introducing the Lagrangian function can obtain:

$$\begin{aligned} L\left(w, b, \alpha\right) &= \frac{\| w \|^2}{2} - \sum_{i=1}^{n} \alpha_i \left[y_i \left(w \cdot x + b\right) - 1\right] \\ &s.t.\ \alpha_i \geq 1, i = 1, 2, \cdots, n \end{aligned} \tag{4}$$

In Eq. 4, α_i refers to the Lagrangian multiplier. The partial derivative of the Lagrangian function can be obtained. According to the extreme conditions:

$$\begin{aligned} \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} y_i \alpha_i = 0 \\ \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} y_i \alpha_i x_i \end{aligned} \tag{5}$$

Substituting equation 5 into 4 can get the dual problem about the original problem:

$$\begin{aligned} &\min_{\alpha} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j \left(x_i \cdot x_j\right) - \sum_{i=1}^{n} \alpha_i \\ &s.t.\ \sum_{i=1}^{n} y_i \alpha_i = 0, \alpha_i y_i \geq 0, i = 1, 2, \cdots, n \end{aligned} \tag{6}$$

Solving Eq. 6 can get the optimal solution $\alpha^* = \left(\alpha_1^*, \cdots, \alpha_n^*\right)^{\mathrm{T}}$. Calculating w^* and b^* can obtain the optimal classification function:

$$f(x) = sgn\left[\sum_{i=1}^{n} \alpha_i^* y_i \left(x_i \cdot x\right) + b^*\right] \tag{7}$$

According to Eq. 7, the optimal classification plane of SVM only depends on the sample point (x_i, y_i) corresponding to $\alpha_i^* > 0$ in the training data. SVM's optimal classification plane only depends on corresponding sample points in the training data, and these

samples just fall on the interval boundary. Usually, the sample point $x_i \in R^n$ at $\alpha_i^* > 0$ is called the support vector. The graph feature extraction of S3VMs is analyzed further.

A training set including labeled data and unlabeled data $\left\{(x_i, y_i)_{i=1}^{l}, (x_j)_{j=l+1}^{l+u}\right\}$ is given. Suppose that $g = (V, E)$ refers to a graph, V represents the node set, and E describes the edge set. The weight matrix W of g is symmetric, which can be described as:

$$W_{ij} = \begin{cases} w(e) = w_{ij} = 0 \ \ if \ \ e = (x_i, x_j) \in E \\ 0 \qquad\qquad\qquad otherwise \end{cases} \tag{8}$$

The weight of the edge $e = (x_i, x_j)$ $w(e) = w_{ij} = w_{ji}$ signifies the similarity between nodes x_i, x_j, which can be calculated using K Nearest Neighbor (KNN) graph. Each vertex in the KNN graph is only connected to its k nearest neighbors under a particular distance (such as Euclidean distance). Notably, a vertex x_i that is the k nearest neighbor of x_j does not necessarily indicate that x_j is the k nearest neighbor of x_i. While constructing the KNN graph, if x_i and x_j are k-nearest neighbors of each other, x_i and x_j will be connected by an edge, and the corresponding weights will be calculated using the Radial Basis Function (RBF) kernel:

$$W_{ij} = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right) \tag{9}$$

In Eq. 9, σ refers to the kernel function that controls the weight reduction speed. When $x_i = x_j$, the weight $w_{ij} = 1$; when $\| x_i - x_j \|$ approaches infinite, the weight $w_{ij} = 0$.

To extract and forecast the obtained brain image features, CNN is introduced into the brain image feature extraction and diagnosis system. The proposed model can minimize the preprocessing workload and directly extract the most expressive features from the original data input without manual feature designation (Ogino et al., 2021). **Figure 3** illustrates how CNN extracts and classifies the features of the graph data in brain image DTs.

Convolutional Neural Network pooling layer performs down-sampling operations on the input feature maps in length and width dimensions. It can reduce the model parameters by down-sampling the input feature data. According to the number of parameters, the data complexity of different brain images can be reduced, diminishing the over-fitting degree and the probability of local minimum. Moreover, the pooling layer can make the model more robust to translation and distortion in the image.

Data in CNN go through multiple convolutional layers and pooling layers; then, they are connected via one or more fully connected layers. All neurons in the current layer are connected to those in the last layer. Usually, this layer depends on two 1D network layers. Local information of the convolutional layer or the pooling layer is grouped together. The activation function used by all neurons is often the Rectified Linear Unit (ReLU).

AlexNet (Tanveer et al., 2019), a deep CNN model, is selected considering its multiple network layers and stronger learning ability to reduce the calculation amount and enhance CNN's generalization performance. Furthermore, the functional layer of AlexNet's convolutional layer is improved. The operation of "local normalization before pooling" is advanced to "pooling before local normalization." This improvement brings two benefits. First, the generalization ability of AlexNet can be enhanced while the over-fitting can be weakened, which greatly shortens the training time. Second, overlapping pooling before local normalization can reserve more data and weaken redundant information during pooling and accelerate the convergence rate of the brain image DTs diagnosis and forecasting model during training, highlighting its superiority over other pooling approaches.

Brain Image Fusion DTs Diagnosis and Forecasting Model Based on S3VMs and Improved AlexNet

Digital Twins (Sun et al., 2020) can forecast and analyze data of the physical space in the corresponding virtual space regarding the smart city development in physical spaces. To extract the data features from brain images, AlexNet, a deep CNN model, is selected considering its multiple network layers and stronger learning ability. Furthermore, the functional layer of AlexNet's convolutional layer is improved. A brain image DTs diagnosis and forecasting model is designed based on S3VMs and improved AlexNet while considering the model's security performance, as demonstrated in **Figure 4**.

In this diagnosis and forecasting model, different penalty coefficients are provided considering the different contribution degrees of the labeled and unlabeled data samples in the brain

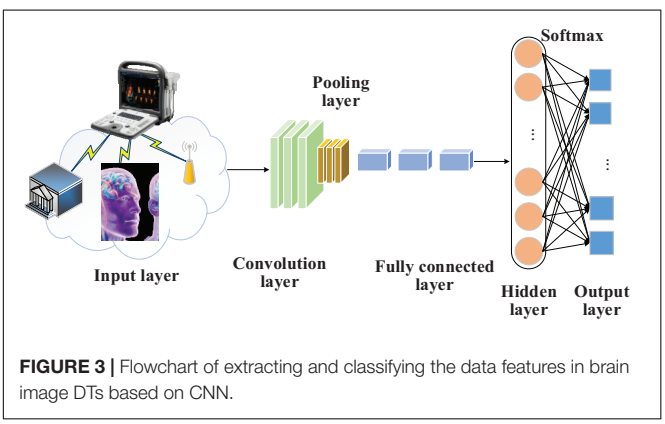


FIGURE 3 | Flowchart of extracting and classifying the data features in brain image DTs based on CNN.

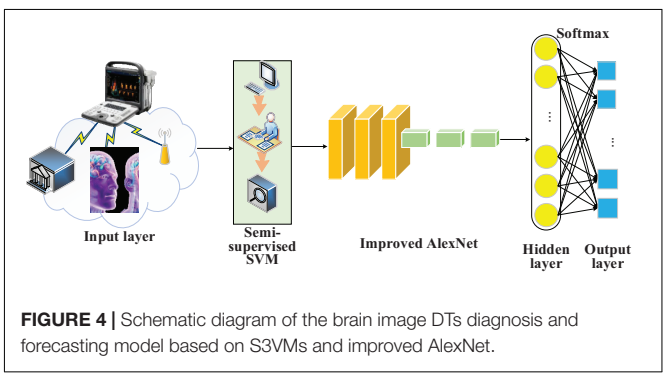


FIGURE 4 | Schematic diagram of the brain image DTs diagnosis and forecasting model based on S3VMs and improved AlexNet.

images to the hinge loss. The objective function can be written as the following equation:

$$\begin{aligned}\min_{w,b,s,y_u}\ & \frac{1}{2}w^T w+\lambda_1\sum_{i=1}^{n}u_{ii}\left(1-\left(w^T x^i+b\right)y_i\right)\\ & +\lambda_2 tr\left(w^T X L_s X^T w\right)+\lambda_3\parallel s\parallel_2^2+\lambda_4 tr\left(F^T L_s F\right)\\ s.t.\quad & \forall i, s_i 1\geq 0, F\in R^{n\times c}, F^T F=I\end{aligned} \tag{10}$$

In Eq. 9, s signifies a similarity matrix between samples, $L_s=D_s-\frac{s^T+s}{2}$, and $D_s\in R^{n\times n}$ refers to a diagonal element, the diagonal matrix of $\sum_j(s_{ij}+s_{ji})/2\leftarrow y_u^{(t)}t=t+1$. The above objective function can be solved by iterative optimization. The optimization process is described in **Figure 5**.

In this algorithm flow, $\lambda_3=\lambda_4$ and s are initialized:

$$\lambda_3=\lambda_4=\frac{1}{n}\sum_{i=2}^{n}\left(\frac{k}{2}d_{i,k+1}^x-\frac{1}{2}\sum_{j=1}^{k}d_{ij}^x\right) \tag{11}$$

$$s_{ij}=\begin{cases}0, & x^j\notin N_k\left(x^i\right)\\ \dfrac{d_{i,k+1}^x-d_{ij}^x}{kd_{i,k+1}^x-\sum_{j=1}^{k}d_{ij}^x}, & x^j\in N_k\left(x^i\right)\end{cases} \tag{12}$$

1	**start**
2	Input: $X\in R^{d\times n}, y\in R^{n\times 1}, U\in R^{n\times n}$
3	Parameters: λ_1,λ_2,k
4	Output: $w\in R^{n\times 1}, b$
5	Initialization: $w^{(0)}=1, b^{(0)}=0, t=1,$ $s^{(0)}\leftarrow$ based on formula (11), $\lambda_3=\lambda_4\leftarrow$ based on formula (10)
6	**While** non convergence **do**
7	Update $L_s{}^t=D_s-\left(s^{(t-1)T}+s^{(t-1)}\right)/2$
8	$D_s\leftarrow$ diagonal matrix
9	$\sum_j(s_{ij}+s_{ji})/2\leftarrow$ the i-th diagonal element
10	Update $F^{(t)}$ based on formula (12)
11	Update $w^{(t)}, b^{(t)}$
12	Update $s^{(t)}$ based on formula (17)
13	Update $y_u^{(t)}$
14	$t=t+1$
15	End while
16	Return w, b
17	**end**

FIGURE 5 | Algorithm flow of objective function's iterative optimization.

Here, $d_{ij}^x=\parallel x^i-x^j\parallel_2^2$. The smallest top k d_{ij}^x are found by sorting in increasing order, and $N_k\left(x^i\right)$ refers to the k samples nearest to x^i. $F^{(t)}$ is updated as the following equation:

$$\begin{aligned}\min_F\ & tr\left(F^T L_s F\right)\\ & s.t. F\in R^{n\times c}, F^T F=I\end{aligned} \tag{13}$$

Eigenvectors corresponding to the smallest c eigenvalues of the Laplacian matrix L_s can form F.

Solving $s^{(t)}$ only needs to solve Eq. 13:

$$\begin{aligned}\min_{s_i 1=1, s_i\geq 0}\ & \sum_{i=1}^{n}\left(\lambda_2\parallel w^T x^i-w^T x^j\parallel_2^2 s_{ij}+2\lambda_3 s_{ij}^2\right)\\ & +\sum_{i=1}^{n}\lambda_4\parallel f_i-f_j\parallel_2^2 s_{ij}\end{aligned} \tag{14}$$

Features of brain images are extracted. The t-th feature map $y_t^l(i,j)$ of the l-th convolutional layer is sampled using overlapping pooling:

$$\begin{aligned}a_t^l(i,j)=\max\{y_t^l(i,j), i_s\leq i\leq i_s+w_c-1,\}\\ j_s\leq j\leq j_s+w_c-1\end{aligned} \tag{15}$$

In Eq. 14, s is the pooling movement step size, w_c refers to the width of the pooling area, and $w_c>s$.

A local normalization layer is added after the first and second pooling layers of AlexNet to standardize the feature map $c_t^l(i,j)$:

$$c_t^l(i,j)=a_t^l(i,j)/\left(k+\alpha\sum_{\max(0,t-m/2)}^{\min(N-1,t+m/2)}\left(a_t^l(i,j)\right)^2\right)^{\beta} \tag{16}$$

In Eq. 15, k, α, β, m are all hyperparameters valuing 2, 0.78, 10^{-4}, and 7, respectively, and N is the total number of convolution kernels in the l-th convolutional layer. To prevent "gradient dispersion" (Shen et al., 2019), the activation function takes Rectified Linear Unit (ReLU) to activate the convolution output $S_t^l(i,j)$:

$$y_t^l(i,j)=f\left(S_t^l(i,j)\right)=\max\left\{0, S_t^l(i,j)\right\} \tag{17}$$

In Eq. 16, $f(\ )$ represents ReLU. To prevent over-fitting in the fully connected layer, the dropout parameter is set to 0.5. All the feature maps when l values 5 in Eq. 12 are reconstructed into a high-dimensional single-layered neuron structure C^5; thus, the input Z_i^6 of the i-th neuron in the sixth fully connected layer is:

$$Z_i^6=W_i^6 C^5+b_i^6 \tag{18}$$

In Eq. 17, W_i^6 and b_i^6 are the weight and bias of the i-th neuron in the sixth fully connected layer, respectively.

While improving the generalization ability, the neurons C^l of the sixth and seventh fully connected layers are discarded and output, $r_j^l\sim bernoulli(dp)$, $\tilde{C}^l=r^l C^l$; in this regard, the i-th neuron's input in the seventh and eighth fully connected layers Z_i^{l+1} is $W_i^{l+1}\tilde{C}^l+b_i^{l+1}$, where the i-th neuron's input in the

sixth and seventh fully connected layers C_i^l is $f\left(Z_i^l\right)$, namely $\max\left\{0, Z_i^l\right\}$. Finally, the input q^i of the i-th neuron in the eighth fully connected layer can be obtained:

$$q^i = soft\max(Z_i^8) = \frac{e^{Z_i^8}}{\sum_{j=1}^{12} e^{Z_i^8}} \tag{19}$$

The cross-entropy loss function suitable for classification is taken as the model's error function, and the equation is:

$$Loss = \sum_{i=1}^{K} y_i \cdot \log\left(p_i\right) \tag{20}$$

$$p_i = \frac{\exp\left(\tilde{y}_i\right)}{\sum_{i=1}^{K} \exp\left(\tilde{y}_j\right)} \tag{21}$$

In Eqs. 19 and 20, K denotes the number of categories, y_i describes the true category distribution of the sample, $\tilde{y}_i$ describes the network output, and p_i represents the classification result after the SoftMax classifier. SoftMax's input is an N-dimensional real number vector, denoted as x. Its equation is:

$$\xi(x)_i = \frac{e^{x_i}}{\sum_{n=1}^{N} e^{x_i}}, i = 1, 2, ..., N \tag{22}$$

Essentially, SoftMax can map an N-dimensional arbitrary real number vector to an N-dimensional vector whose values all fall in the range of (0,1), thereby normalizing the vector. To reduce the computational complexity, the output data volume is reduced to 2^8 through μ companding conversion, that is, $\mu = 255$, thereby improving the model's forecasting efficiency.

$$f(x_t) = sign(x_t)\frac{\ln(1+\mu|x_t|)}{\ln(1+\mu)}, |x_t| < 1 \tag{23}$$

The proposed model is trained through learning rate updating using the polynomial decay approach (Poly) (Zhao et al., 2020). The equation is:

$$init_lr \times \left(1 - \frac{epoch}{\max_epoch}\right)^{power} \tag{24}$$

In Eq. 23, the initial learning rate *init_lr* is 0.0005 (or $5e^{-4}$), and power is set to 0.9.

The Weighted Cross-Entropy (WCE) is accepted as a cost function to optimize model training. Suppose that $z_k(x, \theta)$ describes the unnormalized logarithmic probability of pixel x in the k-th category under the given network parameter θ. In that case, the SoftMax function $p_k(x, \theta)$ is defined as:

$$p_k(x, \theta) = \frac{\exp\{z_k(x, \theta)\}}{\sum_{k'}^{K} \exp\{z_{k'}(x, \theta)\}} \tag{25}$$

In Eq. 24, K represents the total number of image categories. During forecasting, once Eq. 24 reaches the maximum, pixel x will be labeled as the k-th category, namely $k^* = \arg\max\{P_k(x, \theta)\}$. A semantic segmentation task needs to sum the pixel data loss in each input mini-batch. Here, N denotes the total number of pixels in the training batch of image data, y_i refers to the real semantic annotation of the pixel x_i, and $p_k(x_i, \theta)$ describes the forecasted probability of pixel x_i belonging to the k-th semantic category, that is, the log-normalized probability, abbreviated as p_{ik}. Hence, the training process aims to find the optimal network parameter θ^* by minimizing the WCE loss function $\ell(x, \theta)$, denoted as $\theta^* = \min_{\theta} \ell(x, \theta)$. Training samples with unbalanced categories in brain images usually make the network notice some easily distinguishable categories, resulting in poor recognition on some more difficult samples. In this regard, the Online Hard Example Mining (OHEM) strategy (Ghafoori et al., 2018) is adopted to optimize the network training process. The improved loss function is:

$$\ell(x, \theta) = -\frac{1}{\sum_{i=1}^{N}\sum_{k=1}^{K} \delta\left(y_i = k, p_{ik} < \eta\right)} \sum_{i=1}^{N}\sum_{k=1}^{K} \delta\left(y_i = k, p_{ik} < \eta\right) \log p_{ik} \tag{26}$$

In Eq. 25, $\eta \in (0, 1]$ refers to the predefined threshold, and $\delta(\cdot)$ describes the symbolic function, which will take 1 if the condition is met and 0 otherwise. The weighted loss function for brain image fusion is defined as:

$$\ell(x, \theta) = -\sum_{i=1}^{N}\sum_{k=1}^{K} w_{ik} q_{ik} \log p_{ik} \tag{27}$$

In Eq. 26, $q_{ik} = q(y_i = k|x_i)$ denotes the true label distribution of the k-th category of the pixel x_i, w_{ik} refers to the weighting coefficient. The following strategy is employed during training:

$$w_{ik} = \frac{1}{\ln(c + p_{ik})} \tag{28}$$

In Eq. 27, c is an additional hyperparameter, set to 1.10 based on experience during simulation experiments.

Simulation Experiment

In the present work, MATLAB is adopted for simulation to validate the performance of the proposed model. Brain image data in experiments come from MRI records of the Brain Tumor Department of a Hospital. It contains 20 groups of clinical data, and each group of data includes four kinds of multi sequence MRI images: T1, T2, Tlc, and FLAIR. The selected image data are processed by registration, skull peeling, contrast enhancement and other operations. Each case of MRI image contains the standard segmentation results given by artificial experts. Data collected are divided into a training dataset and a test dataset in 7:3; the ratio of each data type in the two datasets shall be consistent. Hyperparameters of AlexNet are set as follows: 120 iterations, 2,000 seconds of simulation, and 128 Batch Size. Some state-of-art models are included for performance comparison, including LSTM (Nguyen et al., 2021), CNN, RNN (Minerva et al., 2021), AlexNet, and MLP (Dong et al., 2019). Experimental environment configuration includes software and hardware. As

for software, the operating system is Linux 64bit, the Python version is 3.6.1, and the development platform is PyCharm. As for hardware, the Central Processing Unit (CPU) is Intel Core i7-7700@4.2 GHz 8 Cores, the internal memory is Kingston DDR4 2400 MHz 16G, and the Graphics Processing Unit (GPU) is NVIDIA GeForce 1060 8G.

The similarity between the experimental segmentation result and the expert segmentation result can measure the model's segmentation quality (Dai et al., 2020). Common similarity metrics are the Jaccard coefficient and Dice Similarity Coefficient (DSC). DSC indicates the similarity degree between the segmentation result of the model and the label labeled by the expert. The higher the DSC, the higher the segmentation similarity. Positive Predictive Value (PPV) represents the proportion of correct brain tumor points in the experimental segmentation to the tumor points in the segmentation result. Sensitivity describes the proportion of the correct tumor points in the experimental segmentation to the true value of the tumor points. The above indicators are adopted to assess the proposed brain image segmentation model. The Jaccard coefficient, DSC, PPV, and Sensitivity are defined as follows:

$$Jaccard(T, P) = \frac{|T \wedge P|}{|T \vee P|} \tag{29}$$

$$DSC(T, P) = \frac{2\,|T \wedge P|}{|T| + |P|} \tag{30}$$

$$PPV = \frac{|T \wedge P|}{P} \tag{31}$$

$$Sensitivity = \frac{|T \wedge P|}{T} \tag{32}$$

In Eqs. 28–31), T represents the expert segmentation result, and P denotes the segmentation result of the proposed model based on S3VMs and the improved AlexNet.

RESULTS AND DISCUSSION

Forecasting Performance Analysis

To explore the forecasting performance of the proposed model, LSTM, CNN, RNN, AlexNet, and MLP are included for comparison. Values of Accuracy, Precision, Recall, and F1 of these models are compared through simulation experiments, and the results are illustrated in **Figure 6**. Furthermore, time durations required for training and test and errors are compared, as in **Figures 7**, **8**.

As shown in **Figure 6**, the proposed model can provide a recognition accuracy of 92.52%, at least an improvement of 2.76% than other models. The Precision, Recall, and F1 of the proposed model are the highest, at least 3.35% higher than other models. Hence, the proposed brain image DTs diagnosis and forecasting model based on S3VMs and improved AlexNet can provide excellent recognition and forecasting accuracy.

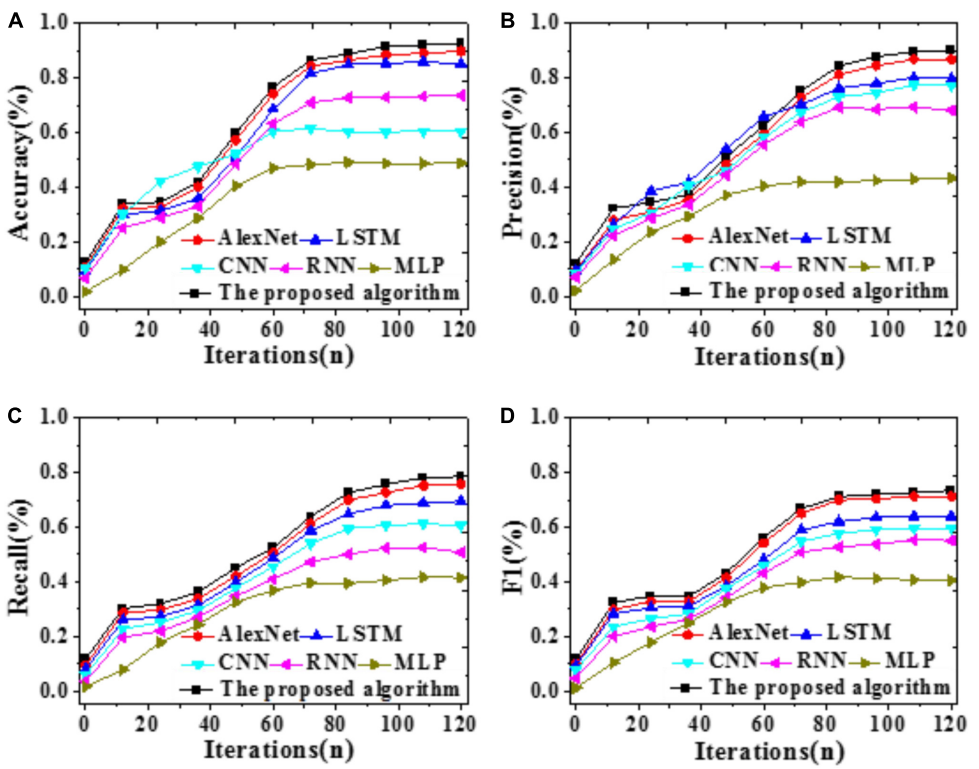


FIGURE 6 | Recognition accuracy of different models with iterations **(A)** Accuracy; **(B)** Precision; **(C)** Recall; **(D)** F1.

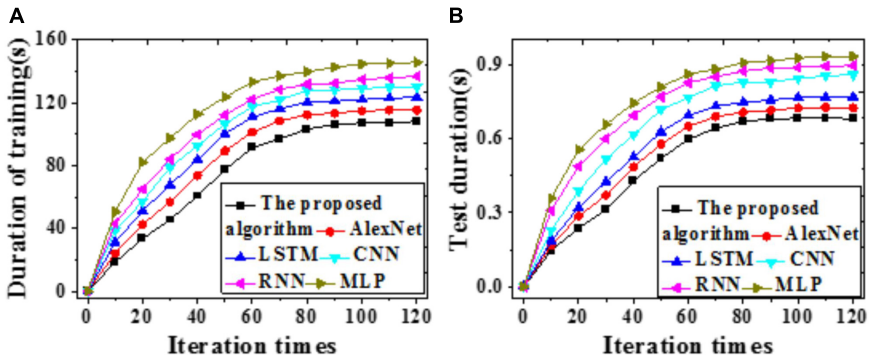


FIGURE 7 | Time duration required by different models **(A)** Training duration; **(B)** Test duration.

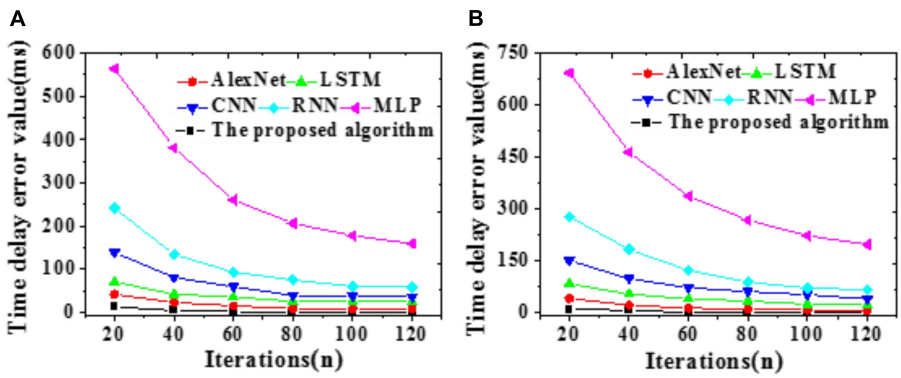


FIGURE 8 | Time delay errors of different models **(A)** Training set; **(B)** Test set.

According to **Figures 7**, **8**, the time required by all models first increases then stabilizes with iterations. Compared with LSTM, AlexNet, CNN, RNN, and MLP, the proposed model requires remarkably less time for training and test. A possible reason is that the improved AlexNet provides enhanced generalization ability while accelerating the convergence velocity of the model training process. Therefore, the proposed model can achieve higher forecasting accuracy more quickly.

In both the training and test sets, errors reduce gradually with iterations. MLP provides the longest time delay, 564.15 ms and 693.06 ms, respectively. In contrast, the proposed model provides a time delay approaching zero, the smallest among all tested models. Comparisons of RMSE and MAE values are summarized in **Tables 1**, **2** below:

According to **Tables 1**, **2**, the RMSE and MAE of the proposed model are 4.91 and 5.59%, respectively, significantly lower than other models. Hence, it can provide stronger robustness and better accuracy for brain image diagnosis and forecasting.

Comparing Assessment Indicators of Brain Image Segmentation and Fusion

The Jaccard coefficients, DSCs, PPVs, and Sensitivity of the proposed model, LSTM, CNN, RNN, AlexNet, and MLP are compared. The results are demonstrated in **Figure 9**.

TABLE 1 | RMSE (%) changes of each model with iterations.

	Iterations		
	1.00	**60.00**	**120.00**
The proposed model	4.85	5.09	4.78
AlexNet	5.61	5.79	5.31
LSTM	6.43	6.11	6.01
CNN	7.28	6.51	6.45
RNN	7.99	6.87	7.04
MLP	8.42	7.33	7.48

TABLE 2 | MAE (%) changes of each model with iterations.

	Iterations		
	1	**60**	**120**
The proposed model	5.67	5.45	5.66
AlexNet	6.59	6.98	6.91
LSTM	8.92	8.26	8.44
CNN	9.93	9.11	8.92
RNN	10.53	9.59	9.48
MLP	11.22	10.19	10.01

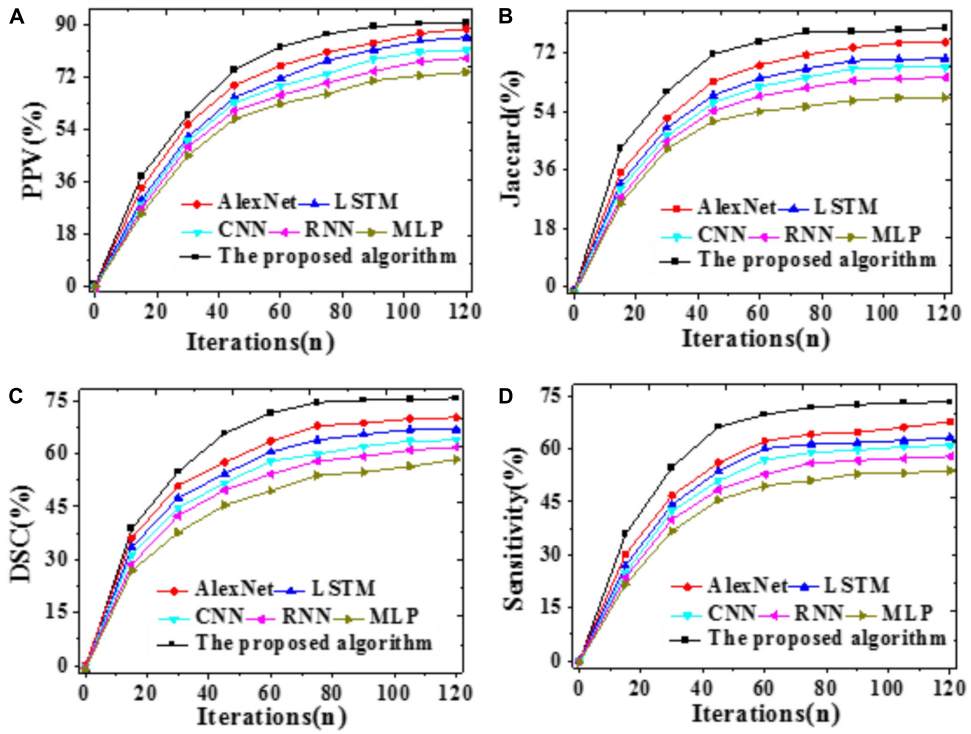


FIGURE 9 | Brain image assessment indicators of different models with iterations **(A)** The Jaccard coefficient; **(B)** DSC; **(C)** PPV; **(D)** Sensitivity.

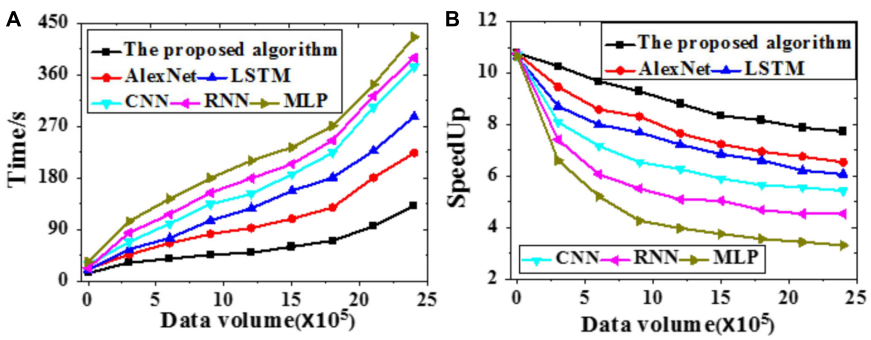


FIGURE 10 | Time required and speedup indicator of different models under different data volumes **(A)** Time required; **(B)** Speedup indicator.

The higher the indicator value, the more accurate the segmentation result. As shown in **Figure 9**, the proposed model can provide a 79.55% Jaccard coefficient, a 90.43% PPV, a 73.09% Sensitivity, and a 75.58% DSC, remarkably better than other models. Hence, the proposed model outperforms other state-of-art models for MRI brain tumor image segmentation.

Acceleration Efficiency Analysis

The acceleration efficiency of the proposed model is analyzed through simulation experiments. The time required and speedup indicator of different models under different data volumes are presented in **Figure 10**.

The proposed model is less sensitive to brain image data growth than other models. Hence, it is suitable to process massive data. Simultaneously, the larger the data volume, the higher the speedup indicator. To sum up, the proposed model can provide better data recognition and segmentation results than other models.

A diagnosis and prediction model of brain image fusion digital twins based on semi supervised SVM and improved AlexNet is proposed to improve the accuracy of traditional CNN segmentation of MRI brain tumor images. The model mainly includes the preprocessing of brain image data based on semi supervised SVM. Then, the improved AlexNet network algorithm is used to extract and analyze the brain image features. Finally, through the comparative test and verification of simulation experiments, the model algorithm constructed can acquire more accurate segmentation results and prediction

accuracy, and can effectively solve a series of problems in the process of MRI brain tumor image fusion, such as under-segmentation and over-segmentation, and meet the needs of clinical pathological diagnosis.

CONCLUSION

Brain tumor images have complicated edge structures, artifacts, offset fields, and other defects that affect image segmentation. Extracting features from multi-sequence MRI brain tumor images becomes particularly vital. Both unlabeled and labeled data are utilized to construct the S3VMs regarding the massive number of unlabeled brain image data. Afterward, AlexNet is improved, and a brain image fusion DTs diagnosis and forecasting model is built based on S3VMs and improved AlexNet. Simulation experiments are performed to analyze the proposed model's performance. Results demonstrate a 92.52% feature recognition and extraction accuracy, a 79.55% Jaccard coefficient, a 90.43% PPV, a 73.09% Sensitivity, and a 75.58% DSC of the proposed model, providing an experimental basis for brain image feature recognition and digital diagnosis. There are some weaknesses in the present work. The constructed model does not grade brain tumors. In the future, the texture feature, shape feature, position information, and volume of the multi-sequence MRI image will be extracted based on brain tumor segmentation results before being classified by ML and DL algorithms, achieving the key steps of precise brain tumor treatment and truly meeting the clinical needs, which will provide remarkable significance for the subsequent clinical diagnosis and treatment of brain tumors.

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

ZW was responsible for the writing and implementation of the full text. YD was responsible for the implementation of the experiment. ZY was responsible for the collection of data and the writing of related research. HL was responsible for the implementation and data analysis of the experiment. ZL was responsible for the design of the experiment and the inspection of the results.

ACKNOWLEDGMENTS

We thank the editorial office of Frontiers in Neuroscience for providing us with important information on our sample of articles, and Sarah Mannion de Hernandez for proofreading.

REFERENCES

Aheleroff, S., Polzer, J., Huang, H., Zhu, Z., Tomzik, D., Lu, Y., et al. (2020). *Smart Manufacturing Based on Digital Twin Technologies. Industry 4.0: Challenges, Trends, and Solutions in Management and Engineering*. (Boca Raton, FL: CRC Press), 77–82.

Ahmed, M. R., Zhang, Y., Liu, Y., and Liao, H. (2020). Single volume image generator and deep learning-based ASD classification. *IEEE J. Biomed. Health Inform.* 24, 3044–3054. doi: 10.1109/jbhi.2020.2998603

Angin, P., Anisi, M. H., Goksel, F., Gursoy, C., and Buyukgulcu, A. (2020). AgriLoRa: a digital twin framework for smart agriculture. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable. Appl.* 11, 77–96.

Barricelli, B. R., Casiraghi, E., and Fogli, D. (2019). A survey on digital twin: definitions, characteristics, applications, and design implications. *IEEE Access* 7, 167653–167671. doi: 10.1109/access.2019.2953499

Dai, Y., Zhang, K., Maharjan, S., and Zhang, Y. (2020). Deep reinforcement learning for stochastic computation offloading in digital twin networks. *IEEE Trans. Ind. Inform.* 17, 4968–4977. doi: 10.1109/tii.2020.3016320

Dong, R., She, C., Hardjawana, W., Li, Y., and Vucetic, B. (2019). Deep learning for hybrid 5G services in mobile edge computing systems: learn from a digital twin. *IEEE Trans. Wirel. Commun.* 18, 4692–4707. doi: 10.1109/twc.2019.2927312

Ghafoori, Z., Erfani, S. M., Rajasegarar, S., Bezdek, J. C., Karunasekera, S., and Leckie, C. (2018). Efficient unsupervised parameter estimation for one-class support vector machines. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 5057–5070. doi: 10.1109/tnnls.2017.2785792

Ghosh, K., Stuke, A., Todorović, M., Jørgensen, P. B., Schmidt, M. N., Vehtari, A., et al. (2019). Deep learning spectroscopy: neural networks for molecular excitation spectra. *Adv. Sci.* 6:1801367. doi: 10.1002/advs.201801367

Gudigar, A., Raghavendra, U., Ciaccio, E. J., Arunkumar, N., Abdulhay, E., and Acharya, U. R. (2019). Automated categorization of multi-class brain abnormalities using decomposition techniques with MRI images: a comparative study. *IEEE Access* 7, 28498–28509. doi: 10.1109/access.2019.2901055

Kang, J. S., Chung, K., and Hong, E. J. (2021). Multimedia knowledge-based bridge health monitoring using digital twin. *Multimed. Tools Appl.* 80, 1–16.

Li, X., Cao, J., Liu, Z., and Luo, X. (2020). Sustainable business model based on digital twin platform network: the inspiration from Haier's case study in China. *Sustainability* 12:936. doi: 10.3390/su12030936

Li, X., He, B., Wang, Z., Zhou, Y., Li, G., and Jiang, R. (2021). Semantic-enhanced digital twin system for robot–environment interaction monitoring. *IEEE Trans. Instrum. Meas.* 70, 1–13. doi: 10.1109/tim.2021.3066542

Liu, L., Huang, W., Liu, B., Shen, L., and Wang, C. (2018). Semisupervised hyperspectral image classification via Laplacian least squares support vector machine in sum space and random sampling. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 11, 4086–4100. doi: 10.1109/jstars.2018.2873051

Liu, Y., Xu, Z., and Li, C. (2018). Distributed online semi-supervised support vector machine. *Inform. Sci.* 466, 236–257. doi: 10.1016/j.ins.2018.07.045

Liu, Y., Zhang, L., Yang, Y., Zhou, L., Ren, L., Wang, F., et al. (2019). A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE Access* 7, 49088–49101. doi: 10.1109/access.2019.2909828

Minerva, R., Awan, F. M., and Crespi, N. (2021). Exploiting digital twin as enablers for Synthetic Sensing. *IEEE Internet Comput.* 25, 1–1. doi: 10.1109/mic.2021.3051674

Nguyen, H. X., Trestian, R., To, D., and Tatipamula, M. (2021). Digital twin for 5G and beyond. *IEEE Commun. Mag.* 59, 10–15. doi: 10.1109/mcom.001.2000343

Ogino, M., Kanoga, S., Ito, S. I., and Mitsukura, Y. (2021). Semi-supervised learning for auditory event-related potential-based brain–computer interface. *IEEE Access* 9, 47008–47023. doi: 10.1109/access.2021.3067337

Pan, Y. H., Qu, T., Wu, N. Q., Khalgui, M., and Huang, G. Q. (2021). Digital twin based real-time production logistics synchronization system in a multi-level computing architecture. *J. Manufact. Syst.* 58, 246–260. doi: 10.1016/j.jmsy.2020.10.015

Qi, Q., and Tao, F. (2018). Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *IEEE Access* 6, 3585–3593. doi: 10.1109/access.2018.2793265

Shao, Y., and Chou, K. C. (2020). pLoc_Deep-mVirus: a CNN model for predicting subcellular localization of virus proteins by deep learning. *Nat. Sci.* 12, 388–399. doi: 10.4236/ns.2020.126033

Shen, M., Tang, X., Zhu, L., Du, X., and Guizani, M. (2019). Privacy-preserving support vector machine training over blockchain-based encrypted IoT data in smart cities. *IEEE Internet Things J.* 6, 7702–7712. doi: 10.1109/jiot.2019.2901840

Sun, J., Tian, Z., Fu, Y., Geng, J., and Liu, C. (2020). Digital twins in human understanding: a deep learning-based method to recognize personality traits. *Int. J. Comput. Integr. Manufact.* 33, 1–14. doi: 10.1080/0951192x.2020.1757155

Tanveer, M., Sharma, A., and Suganthan, P. N. (2019). General twin support vector machine with pinball loss function. *Inform. Sci.* 494, 311–327. doi: 10.1016/j.ins.2019.04.032

Tobon Vasquez, J. A., Scapaticci, R., Turvani, G., Bellizzi, G., Rodriguez-Duarte, D. O., Joachimowicz, N., et al. (2020). A prototype microwave system for 3D brain stroke imaging. *Sensors* 20, 2607–2613. doi: 10.3390/s20092607

Wang, S., Guo, X., Tie, Y., Lee, I., Qi, L., and Guan, L. (2018). Graph-based safe support vector machine for multiple classes. *IEEE Access* 6, 28097–28107. doi: 10.1109/access.2018.2839187

Wang, Z., Yao, L., Cai, Y., and Zhang, J. (2020). Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis. *Renew. Energy* 155, 1312–1327. doi: 10.1016/j.renene.2020.04.041

Wu, Y., Luo, Y., Chaudhari, G., Rivenson, Y., Calis, A., De Haan, K., et al. (2019). Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci. Appl.* 8, 1–7. doi: 10.1117/1.jatis.3.4.049002

Yan, H., Wan, J., Zhang, C., Tang, S., Hua, Q., and Wang, Z. (2018). Industrial big data analytics for prediction of remaining useful life based on deep learning. *IEEE Access* 6, 17190–17197. doi: 10.1109/access.2018.2809681

Zavrak, S., and İskefiyeli, M. (2020). Anomaly-based intrusion detection from network flow features using variational autoencoder. *IEEE Access* 8, 108346–108358. doi: 10.1109/access.2020.3001350

Zhang, H., Li, Y., Lv, Z., Sangaiah, A. K., and Huang, T. (2020). A real-time and ubiquitous network attack detection based on deep belief network and support vector machine. *IEEE CAA J. Automatica Sinica* 7, 790–799. doi: 10.1109/jas.2020.1003099

Zhao, H., Zheng, J., Deng, W., and Song, Y. (2020). Semi-supervised broad learning system based on manifold regularization and broad network. *IEEE Trans. Circuits Syst. I Regul. Pap.* 67, 983–994. doi: 10.1109/tcsi.2019.2959886

Zheng, Y., Wang, S., Li, Q., and Li, B. (2020). Fringe projection profilometry by conducting deep learning from its digital twin. *Opt. Express* 28, 36568–36583. doi: 10.1364/oe.410428

Zhou, M., Yan, J., and Feng, D. (2019). Digital twin framework and its application to power grid online analysis. *CSEE J. Power Energy Syst.* 5, 391–398.

Conflict of Interest: YD was employed by Qingdao Haily Measuring Technologies 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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 30 July 2021
doi: 10.3389/fnins.2021.714318

Fuzzy System Based Medical Image Processing for Brain Disease Prediction

Mandong Hu[1], Yi Zhong[1], Shuxuan Xie[1], Haibin Lv[2] and Zhihan Lv[1]*

[1] College of Computer Science and Technology, Qingdao University, Qingdao, China, [2] North China Sea Offshore Engineering Survey Institute, Ministry of Natural Resources North Sea Bureau, Qingdao, China

The present work aims to explore the performance of fuzzy system-based medical image processing for predicting the brain disease. The imaging mechanism of NMR (Nuclear Magnetic Resonance) and the complexity of human brain tissues cause the brain MRI (Magnetic Resonance Imaging) images to present varying degrees of noise, weak boundaries, and artifacts. Hence, improvements are made over the fuzzy clustering algorithm. A brain image processing and brain disease diagnosis prediction model is designed based on improved fuzzy clustering and HPU-Net (Hybrid Pyramid U-Net Model for Brain Tumor Segmentation) to ensure the model safety performance. Brain MRI images collected from a Hospital, are employed in simulation experiments to validate the performance of the proposed algorithm. Moreover, CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), FCM (Fuzzy C-Means), LDCFCM (Local Density Clustering Fuzzy C-Means), and AFCM (Adaptive Fuzzy C-Means) are included in simulation experiments for performance comparison. Results demonstrate that the proposed algorithm has more nodes, lower energy consumption, and more stable changes than other models under the same conditions. Regarding the overall network performance, the proposed algorithm can complete the data transmission tasks the fastest, basically maintaining at about 4.5 s on average, which performs remarkably better than other models. A further prediction performance analysis reveals that the proposed algorithm provides the highest prediction accuracy for the Whole Tumor under DSC (Dice Similarity Coefficient), reaching 0.936. Besides, its Jaccard coefficient is 0.845, proving its superior segmentation accuracy over other models. In a word, the proposed algorithm can provide higher accuracy, a more apparent denoising effect, and the best segmentation and recognition effect than other models while ensuring energy consumption. The results can provide an experimental basis for the feature recognition and predictive diagnosis of brain images.

Keywords: fuzzy system, brain image, digital twins, image segmentation, fuzzy clustering, HPU-Net

OPEN ACCESS

Edited by:
Kaijian Xia,
Changshu No. 1 People's Hospital, China

Reviewed by:
Tongguang Ni,
Changzhou University, China
Changjun Zhou,
Zhejiang Normal University, China

****Correspondence:***
Mandong Hu
mandonghu@126.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *25 May 2021*
Accepted: *12 July 2021*
Published: *30 July 2021*

Citation:
Hu M, Zhong Y, Xie S, Lv H and Lv Z (2021) Fuzzy System Based Medical Image Processing for Brain Disease Prediction. Front. Neurosci. 15:714318. doi: 10.3389/fnins.2021.714318

INTRODUCTION

Now that economy develops unprecedentedly, health has become the top priority that affects people's living standards. As a common disease that affects human health, brain diseases are classified based on their severity. Brain tumors are undoubtedly the brain diseases that pose the greatest threat to human health. Brain tissue is a functional organization with complicated

structure and irregular shape. The accurate segmentation of brain images provides rich and valuable information for clinical surgeries. Nevertheless, medical images of brain tissues are usually susceptible to interference, such as noises, uneven grayscale, local volume effects, and artifacts (Jiang et al., 2019; Senthilkumar and Gnanamurthy, 2019). In the meantime, there are great difficulties in recognizing and segmenting brain images due to the weak contrast in the image edges and the complicated brain tissue structure.

AI (Artificial Intelligence) technologies, such as DL (Deep Learning), BD (Big Data), and IoT (Internet of Things), have been universally accepted in various fields; in that case, intelligent development in the medical field gets increasingly vital for monitoring people's health status. Correct processing of medical images interfered by strong noises and geometric distortions is significant in clinical diagnosis and treatment (Zheng et al., 2020). DL approaches have made significant progress in image resolution recognition and segmentation; nevertheless, real medical images with low resolution are complicated and unknown. Influences of blurring factors, such as weakening of the image boundary and uneven grayscale, are extremely important. Since DNN (Deep Neural Network) does not consider the blurring while training the data, directly applying medical images processed by DL models to real-life settings often results in unideal reconstruction effects and even apparent blurs and artifacts (Pham et al., 2018; Mishro et al., 2020). The fuzzy theory breaks the traditional "one or the other" view, making the image information more thoroughly applied. Meanwhile, the fuzzy system, an unsupervised classification model, does not require manual intervention, whose segmentation process is completely automatic (Li et al., 2020). The uncertainty and ambiguity in the image can be processed well. Thus, correct processing of brain images can have rather practical significance for diagnosing and treating the brain diseases.

The imaging mechanism of NMR (Nuclear Magnetic Resonance) and the complexity of human brain tissues cause the MRI (Magnetic Resonance Imaging) images to present varying degrees of noise, weak boundaries, and artifacts. Therefore, brain imaging processing for brain disease prediction provides extremely significant value for reducing the risks of clinical treatment. The innovative points are the improvements over fuzzy clustering regarding the noises, weak boundaries, and artifacts in brain images, and the brain image processing and brain disease diagnosis prediction model designed based on HPU-Net (Hybrid Pyramid U-Net Model for Brain Tumor Segmentation) and improved fuzzy clustering, to provide some theoretical support for feature recognition and prediction diagnosis based on brain images.

RECENT WORKS

Existing Applications of Fuzzy System

The traditional "one or the other" idea in various fields cannot explain some actual situations. In contrast, the fuzzy logic system can break this traditional idea and is very useful in solving problems in actual applications. Many scholars have explored fuzzy systems. Feng and Chen (2018) combined TS (Takagi-Sugeno) fuzzy system with fuzzy BLS (Broad Learning System) and proposed a new neuro-fuzzy model and fuzzy BLS. In the meantime, they combined the defuzzification output of each fuzzy subsystem with the output of the enhancement layer to get the model output. Additionally, they employed K-Means to determine the center of the Gaussian membership function in the antecedent and the number of fuzzy rules. Eventually, they found that the performance of the fuzzy BLS model was better than that of other models (Feng and Chen, 2018). IT2 (Interval Type-2) fuzzy sampling data H∞ had parameter uncertainty, data loss, and transmission delay; to solve these problems, Du et al. (2019) connected the IT2 fuzzy system describing the network control system and the IT2 network sampling controller implemented by the zero-order holder in a closed-loop system. Then, they adopted the input delay approach to convert the closed-loop system into a continuous time-delay system. Simultaneously, they utilized continuous-time Lyapunov-Krasovskii functional theory for stability analysis. Eventually, the effectiveness and superiority of the designed method were validated through four different actual systems (Du et al., 2019). To promote the exchange of fuzzy systems between different programming systems, Xie et al. (2020) imported/exported fuzzy systems for the operability and practicality of the standard. They provided a case study to illustrate the advantages of fuzzy systems (Xie et al., 2020). Harifi et al. (2020) proposed an ANFIS (Adaptive Neuro-Fuzzy Inference System) based on the EPC (Emperor Penguins Colony) algorithm regarding the unresolved fuzzy status in biological systems or physical systems in nature. Then, on the benchmark dataset, they compared the optimized ANFIS with other non-derivative algorithms. Afterward, this algorithm was applied to solve the classic inverted pendulum problem. Results demonstrated that ANFIS based on EPC provided smaller errors and better performance than other algorithms in the training and testing stages (Harifi et al., 2020).

Current Development of Medical Image Processing

Medical image segmentation is a vital tool for current clinical applications. It is also the backbone of many clinical diagnostic approaches, tumor treatments, and computer-integrated surgeries. Many scholars have researched this issue. Bello et al. (2019) practiced an FCN (Fully Convolutional Network) trained prior to anatomical shape to create a time-resolved 3D segmentation for cardiac image sequences obtained from cardiac MRI. Moreover, they verified model's prediction accuracy, and how to utilize high-dimensional medical image data for complex computer vision tasks to effectively predict human survival was elaborated (Bello et al., 2019). Jeyaraj and Nadar (2019) put forward a new blocked DCNN (Deep Convolutional Neural Network) regarding the medical image of oral cancer. This network had a two-layer structure, which could mark and classify Regions of Interest (ROI) in multi-dimensional hyperspectral images. The classification accuracy of this network was compared with another traditional medical image classification algorithm. Results suggested that for complex medical images of oral cancer

diagnosis, the proposed regression-based blocked DCNN could improve the diagnostic quality (Jeyaraj and Nadar, 2019). Renard et al. (2020) put forward an original overview of source variability to better understand DL medical image segmentation associated with challenges and problem reproducibility. In the end, they applied a DL framework to analyze different sources of variability for potential problems properly, and developed and verified an effective segmentation result evaluation system (Renard et al., 2020). Yoo et al. (2020) explored the feasibility of using DL decision tree classifiers to detect COVID-19 (Coronavirus Disease 2019) from CXR (Chest X-ray Radiography) images regarding the massive global outbreak of COVID-19. Compared with other domain models, the proposed DL-based decision tree classifier could pre-screen patients for classification and quick decision-making before the results of RT-PCR (Reverse Transcription Polymerase Chain Reaction) were available (Yoo et al., 2020).

To sum up, AI algorithms, such as DL, have been accepted in medical image processing; however, features of multiple boundary pixels are not continuous. Medical images with no apparent differences in pixel features on both sides of the boundary cannot be effectively processed. The fuzzy system algorithm is just an effective solution to no apparent difference in image boundaries. Thus, the fuzzy system is introduced based on DL algorithms to process medical images, which has great practical value for the accurate identification and diagnosis of brain images.

Fuzzy System-Based Medical Image Processing for Brain Disease Prediction

Medical imaging equipment gets updated continuously as computer technology advances. Medical image segmentation is a vital link to medical image analysis, as well as a prerequisite and key technology for clinical surgeries. Nonetheless, due to individual differences and the complexity of human tissues such as the brain, the segmentation results of medical images, especially brain tumor images, are not satisfactory. As a result, countless researchers in this field worldwide have invested a lot of energies to research this issue.

Demand of Applying Fuzzy System to Brain Image Diagnosis and Prediction

The amount of information in images keeps increasing with modern medical imaging technologies, making it necessary for medical workers to undertake a huge workload during medical image analysis. However, images appear to have low contrast, variability between tissues, the ambiguity between different tissues or between tissues and lesions due to the differences in the imaging principles of medical images and the characteristics of the tissues themselves (Barricelli et al., 2019; Kong et al., 2019). Therefore, the complexity and variety of medical images themselves have caused great obstacles to the accurate processing and analysis of medical images. Hence, using computer technology to assist brain image recognition and brain disease prediction is of vital significance for diagnosing and treating diseases.

Medical brain images have loads of uncertainty during diagnosis and prediction, and their essence lies in the ambiguity. There are three key manifestations. First, the grayscale is fuzzy: brain images are interfered with by lighting conditions and spatial resolution, making the edge of the pixel grayscale between the brain image boundary and the background blurry and overlap with each other. Second, it has a local body effect: due to the influence of equipment factors, voxels on a boundary often contain two substances: boundary and object, making it difficult to accurately describe the relationships among the edges, corners, and regions of the objects in the image. Third, there is uncertainty knowledge. In the case of pathological changes, brain images will have lumps or masses that are not available in normal tissues, bringing great difficulties in constructing the model. Besides, there may be noise interference, offset field effects, and partial volume effects in brain image processing. **Figure 1** describes the problems of brain medical image processing.

Based on the above problems in brain medical image processing, this work uses the boundary information of unlabeled and labeled data in brain medical image as medical domain knowledge to further improve the effect of brain image segmentation and diagnosis. Moreover, the semi-supervised recognition method is introduced and deep learning algorithm is integrated based on semi-supervised learning. In different environments, the human brain MRI image is segmented. Finally, a brain disease medical image segmentation system is designed based on the improved algorithm.

Segmentation Performance of Fuzzy System for Brain Images

The ambiguity of brain images refers to a characteristic between the brain medical image and the background. The uncertain boundary between the brain image and the background makes it difficult to segment such images because the machine will be affected by various factors during imaging. Actually, the most popular approach based on fuzzy theory is the FCM (Fuzzy C-Means) algorithm (Bai et al., 2018; Li et al., 2021). FCM not only avoids problems such as manually setting the threshold in advance but also has great advantages in dealing with the uncertainty and inaccuracy of images.

The evaluation indicators of cluster analysis are often judged based on the distance between the cluster centers, the number of samples in the cluster structure, and the standard deviation vector of the samples in the dataset and the cluster centers (Maqsood et al., 2020; Rahman et al., 2020; Wang Z. et al., 2020). The dataset X is divided into C categories, and C cluster sets X_1, X_2,..., X_C are obtained. If the C cluster sets completely conform to Equation (1), it will be called the hard C division of X:

$$\begin{cases} X_1 \cup X_2 \cup \ldots \cup X_c = X \\ X_i \cap X_k = \emptyset, 1 \le i \ne k \le c \\ X_i \ne \emptyset, X_i \ne X, 1 \le i \le c \end{cases} \quad (1)$$

Fuzzy C division can determine the degree of uncertainty of each sample in the dataset in all classes. It can help to better understand and describe the features in the brain image. Regarding division

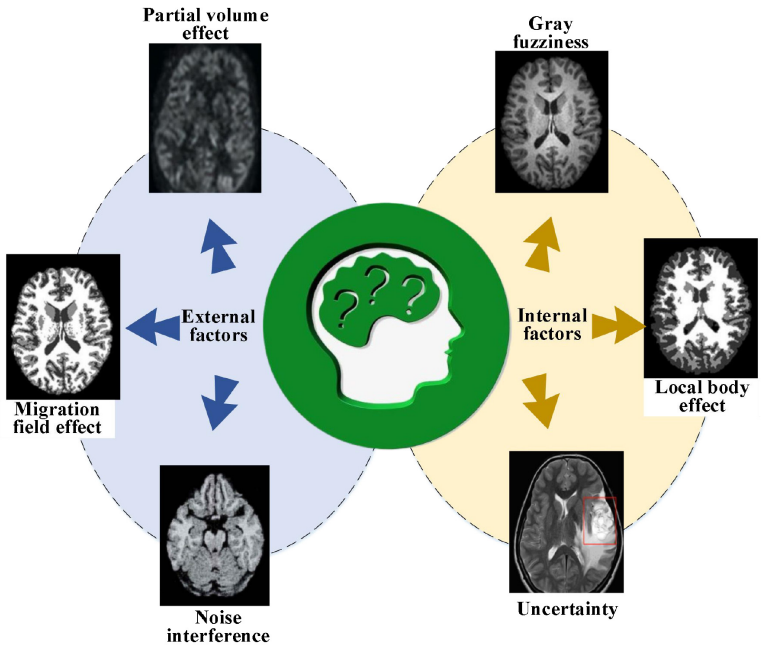


FIGURE 1 | Problems in brain medical image processing.

results, fuzzy division can indicate the periphery of the division, the connection, and the dispersion between different division blocks. Hence, it can mine detailed information about the brain image (Liu et al., 2018). The core task of FCM is to minimize the weighted sum of squares of the distance from the classification sample point to each classification center. The weight is the membership function of each sample point to each classification center, which is the general criterion of the clustering algorithm (Ahmadi, 2020). The objective function of classification is obtained as (2):

$$J_{\min}(U, V) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^{m} d_{ij}^{2} \tag{2}$$

In (2), U refers to the classification matrix, C denotes the number of sample classes, which is determined based on actual situations, u_{ij} describes the corresponding sample membership function, V signifies the classification center vector, $V=[v_1,v_2,v_3,...v_c]$, d_{ij} refers to the distance from the i-th classification center to the j-th sample point, and m is a constant that should be determined before the algorithm runs. The difference between FCM and general clustering algorithm is that it carries out the weighted operation of membership degree. The constraint conditions described in Equation (3) can be obtained from the membership function:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1, \ldots, n \tag{3}$$

According to Equations (2) and (3), the key process of FCM is to find the best combination of U and V so that $J(U,V)$ can achieve the minimum value under constraint conditions. Therefore, Equation (4) can be obtained using the LaGrange multiplier:

$$\begin{aligned} F &= J(U, v_1, \ldots v_c, \lambda_1, \ldots, \lambda_n) \\ &= \textstyle\sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^{m} d_{ij}^{2} + \sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} u_{ij} - 1) \end{aligned} \tag{4}$$

In (4), $\lambda_{j,j=1,2,\ldots,n}$ is the LaGrange operator. The first-order necessary optimization condition for the partial derivative of Equation (4) is:

$$\frac{\partial F}{\partial \lambda} = \sum_{i=1}^{c} u_{ij} - 1 = 0 \tag{5}$$

$$\frac{\partial F}{\partial u_{ij}} = m \left(u_{ij}\right)^{m-1} \left(d_{ij}\right)^{2} - \lambda = 0 \tag{6}$$

Finishing Equation (6) can obtain Equation (7):

$$u_{ij} = \left[\frac{\lambda}{m \left(d_{ij}\right)^{2}} \right]^{1/(m-1)} \tag{7}$$

Substituting Equation (7) into (3) can get Equation (8):

$$\begin{aligned} \textstyle\sum_{i=1}^{c} u_{ij} &= \textstyle\sum_{i=1}^{c} \left(\frac{\lambda}{m}\right)^{1/(m-1)} \left[\frac{1}{\left(d_{ij}\right)^{2}} \right]^{1/(m-1)} \\ &= \left(\frac{\lambda}{m}\right)^{1/(m-1)} \left\{ \textstyle\sum_{i=1}^{c} \left[\frac{1}{\left(d_{ij}\right)^{2}} \right]^{1/(m-1)} \right\} = 1 \end{aligned} \tag{8}$$

Thus, Equation (9) can be obtained:

$$\left(\frac{\lambda}{m}\right)^{1/(m-1)} = \frac{1}{\sum_{i=1}^{c}\left[\frac{1}{\left(d_{ij}\right)^2}\right]^{1/(m-1)}} \tag{9}$$

Substituting Equation (9) into (7) can get Equation (10):

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{d_{ij}}{d_{kj}}\right]^{2/(m-1)}} \tag{10}$$

Because d_{ij} can be zero, two different cases should be considered, respectively. For $\forall j$, two sets I_j and $\tilde{I}_j$ are defined as Equations (11) and (12), respectively:

$$I_j = \left\{i | 1 \le i \le c, d_{ij} = 0\right\} \tag{11}$$

$$\tilde{I}_j = \{1, 2, \cdots, c\} - I_j \tag{12}$$

Hence, the membership function that minimizes $J_{\min(U,V)}$ values:

$$\begin{cases} u_{ij} = \frac{1}{\sum_{k=1}^{c}\left[\frac{d_{ij}}{d_{kj}}\right]^{2/(m-1)}} & I_j = \phi \\ u_{ij} = 0, \forall i \in I_j, \sum_{i \in I_j} u_{ij} = 1 \; I_j \neq \phi \end{cases} \tag{13}$$

Similarly, $J_{\min(U,V)}$ is the smallest cluster center value. Then, the partial derivative of V_i is:

$$\frac{\partial J}{\partial V_i} = \sum_{j=1}^{n}\left(u_{ij}\right)^m \frac{\partial}{\partial V_i}\left[\left(x_j - V_i\right)^T A\left(x_j - V_i\right)\right] = 0 \tag{14}$$

$$\sum_{j=1}^{n}\left(u_{ij}\right)^m \left[-A\left(x_j - V_i\right)\right] = 0 \tag{15}$$

$$\sum_{j=1}^{n}\left(u_{ij}\right)^m \left(x_j - V_i\right) = 0 \tag{16}$$

Eventually, Equation (17) can be obtained:

$$V_i = \frac{\sum_{j=1}^{n}\left(u_{ij}\right)^m x_j}{\sum_{j=1}^{n}\left(u_{ij}\right)^m} \tag{17}$$

According to the number of cluster categories of the brain image data sample set and the prescribed fuzzy weight of the brain image sample, the cluster center and the best classification matrix can be determined by Equations (16) and (17). The objective function can be optimized by iteratively solving its minimum value.

Furthermore, the classic FCM gets improved. Spatial information constraints are introduced to increase the algorithm's resistance to noise in brain images, and the objective function can be obtained:

$$\begin{aligned} &J_{improved}(U, v_1, \ldots v_c, \lambda_1, \ldots, \lambda_n) \\ &= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \left||x_j - V_i\right||^2 + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \sum_{r \in N_j} ||x_r - V_i||^2 \end{aligned} \tag{18}$$

The latter term in Equation (18) is the spatial constraint information, equivalent to the square error about element *i*, and its essence is similar to the mean filtering for point *i*. The noise removal or resistance is performed on the image similar to the nature of spatial filtering so that the algorithm has stronger anti-noise performance. Besides, the edge information of brain images should be enhanced so that the image will not become blurred while denoising. Traditional average filtering and median filtering cannot meet these requirements. Anisotropic diffusion filtering based on partial differential equations, that is, the PM (Perona and Malik) model, can remove the noises in the image while retaining or even enhancing the edge information. Concepts of mass diffusion theory and iterative smoothing are introduced into image processing (Zhang et al., 2020). The PM model is described as:

$$\frac{\partial I}{\partial t} = div(c \cdot \nabla I) \tag{19}$$

In (19), I refers to the original pixel brightness signal, t refers to the diffusion time, ∇ refers to the gradient calculation, *div* refers to the divergence calculation, and c is the diffusion coefficient of the algorithm, namely a function of the image element gradient, and its value range is [0, 1], defined as Equation (20):

$$c = f\left(|\nabla I|\right) \tag{20}$$

The diffusion system determines how the anisotropic algorithm diffuses. An adaptive diffusion control strategy can be used; more diffusion operations can be performed where there is noise in the brain image, and the diffusion can be stopped in the edge area of the image to retain the edge information (Mahata and Sing, 2020). Two diffusion functions commonly used at present are:

$$c_1 = \exp\left(-\left(\frac{\nabla_i I_i}{\lambda}\right)^2\right) \tag{21}$$

In (21), ∇_i represents the gradient of direction *i*, and λ represents the threshold of the gradient that determines how much the edge of the brain image can remain after being diffused.

$$c_2 = \frac{1}{1 + \left(\frac{\nabla_i I_i}{\lambda}\right)^{1+\alpha}} \tag{22}$$

In (22), α demonstrates the parameter of this equation, and $\alpha \in (0 \infty]$. Here, the whole $\frac{\nabla_i I_i}{\lambda}$ is taken as the independent variable of coefficient c. When $\alpha=1$, the characteristic curves of diffusion coefficients c_1 and c_2 are demonstrated in **Figure 2**.

In these two types of commonly used diffusion functions, the gradient threshold plays a vital role during anisotropic diffusion called the diffusion constant (Zhao et al., 2018). The stream function is evaluated to clarify the relationship between diffusion constant λ and diffusion function:

$$\varphi\left(x, t\right) = c\left(x, t\right) \nabla I\left(x, t\right) |\nabla I| < \lambda \tag{23}$$

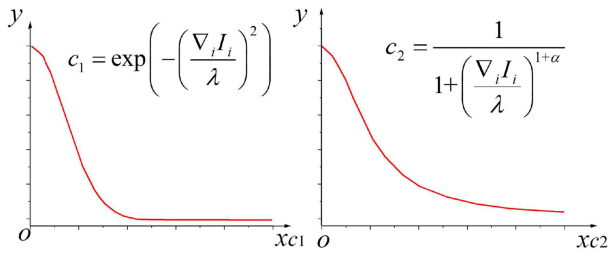


FIGURE 2 | Characteristic curves of diffusion coefficients c_1 and c_2.

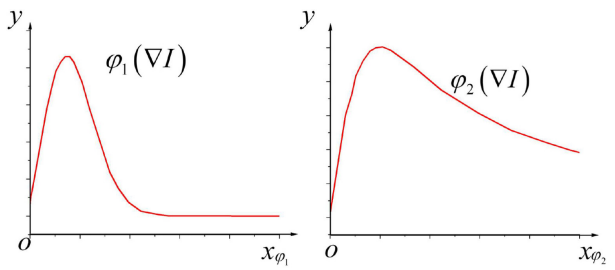


FIGURE 3 | Characteristic curves of stream functions corresponding to diffusion coefficients c_1 and c_2.

$$\frac{\partial I}{\partial t} = div(\varphi(x,t))\,|\nabla I| > \lambda > 1 \tag{24}$$

Besides, $\varphi_1(\nabla I), \varphi_2(\nabla I)$ are defined as the stream functions corresponding to the above two types of diffusion functions; if $\nabla I/\lambda$ serves as the independent variable, characteristic curves of stream functions corresponding to these two diffusion coefficients when $\alpha=1$ are demonstrated in **Figure 3**:

According to **Figure 3**, when the absolute value of the pixel gradient is $|\nabla I|<\lambda$; that is, when $|\nabla I|/\lambda<1$, the currently calculated element is not on the image edge. Simultaneously, as the image gradient increases continuously, the value of the stream function also increases, which increases the grayscale diffusion of the corresponding image elements, thereby removing the noise points. When $|\nabla I|>\lambda$, that is, $|\nabla I|/\lambda>1$, the current element is included in the image edge. At this time, as the image gradient continues to increase, the value of the stream function will continue to decrease instead; consequently, the grayscale diffusion at the corresponding image element will continue to weaken until it becomes zero so that the edge information of the brain image can be well preserved.

Brain Image Processing and Brain Disease Diagnosis Prediction Model Based on Fuzzy Clustering and HPU-Net

The imaging mechanism of NMR and the complexity of human brain tissues make MRI images present various degrees of noise, weak boundaries, and artifacts. Therefore, the fuzzy clustering algorithm is improved to extract the features of data obtained in the brain image. Afterward, a brain image processing and brain disease diagnosis prediction model based on improved fuzzy clustering and HPU-Net is designed while ensuring the safety performance of the model. **Figure 4** illustrates its structure.

In the proposed algorithm, an improved fuzzy clustering algorithm is proposed based on the kernel distance metric regarding the strong noise interference of medical images. This algorithm combines the kernel function, membership constraints, and regularization parameter ρ based on the traditional fuzzy clustering algorithm, making full use of the local spatial information of the image. This makes the algorithm possess better segmentation accuracy and detail retention ability interfered by high-intensity noise, and the image edges after segmentation have better smoothness. This algorithm is recorded as SIPMFCM (Fuzzy Clustering based on Spatial Information Fusion PM) model, described as Equation (25):

$$\begin{aligned} J_{SIPMFCM}(U, v_1, \ldots v_c, \lambda_1, \ldots, \lambda_n) &= 2\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\left(1-||x_j - v_i||\right) \\ &+ \sum_{i=1}^{c}\rho_i\sum_{j=1}^{n} u_{ij}\left(1-u_{ij}^{m-1}\right) + 2\sum_{i=1}^{c}\rho_i\sum_{j=1}^{n}\left(1-||\bar{x}_j - v_i||\right) \end{aligned} \tag{25}$$

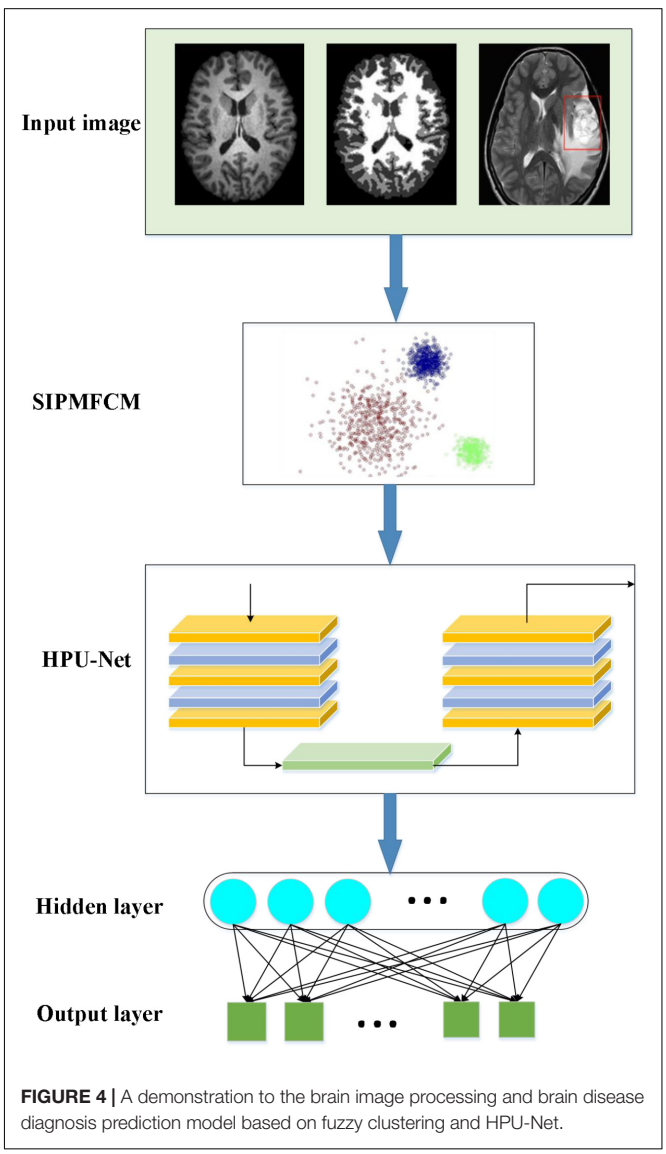


FIGURE 4 | A demonstration to the brain image processing and brain disease diagnosis prediction model based on fuzzy clustering and HPU-Net.

start
Input: p, q, m
Parameters: p=2, q=1, m=2, iteration number t=0, threshold $\varepsilon = 0.001, u^{(0)}$
Output: $u_{ij}^{'}$
Calculate the mean filter image and median filter image $\overline{x}_j$
Calculate The cluster center $v_i^{(t)}$

$$v_i \leftarrow \frac{\sum_{j=1}^{n} u_{ij}^{m}\left(\left(\left\|x_j - v_i\right\|\right)x_j + \rho_i \left\|\overline{x}_j - v_i\right\|\overline{x}_j\right)}{\sum_{i=1}^{c} u_{ij}^{m}\left(\left\|x_j - v_i\right\| + \rho_i \left\|\overline{x}_j - v_i\right\|\right)}$$

Calculation of membership function u_{ij}

$$u_{ij} \leftarrow \frac{1}{\sum_{k=1}^{c}\left(\frac{2\left(1-\left\|x_j - v_i\right\|\right) - \rho_i + 2\rho_i\left(1-\left\|x_j - v_i\right\|\right)}{2\left(1-\left\|x_j - v_k\right\|\right) - \rho_i + 2\rho_i\left(1-\left\|x_j - v_k\right\|\right)}\right)^{1/(m-1)}}$$

if $\left\|u^{(t+1)} - u^{(t)}\right\| < \varepsilon$ or t>100
  Iteration termination
  then update t=t+1
  Return $v_i^{(t)}, u_{ij}$
end if
end

FIGURE 5 | Flowchart of the improved fuzzy clustering algorithm SIPMFCM.

In (25), $\bar{x}$ refers to the mean filtering brain image or median filtering brain image. The first item here is that the traditional FCM expression is transformed from the kernel distance measurement, which maps the samples to a high-dimensional space and increases the difference between the samples. The second membership penalty item makes the fuzzy division more distinct. The third item here is the neighborhood space restriction item, aiming to strengthen the algorithm's robustness to image noises. Eventually, the spatial function is integrated into the membership function u_{ij}:

$$u_{ij}^{'} = \frac{u_{ij}^{p} s_{ij}^{q}}{\sum_{k=1}^{c} u_{ik}^{p} s_{ik}^{q}} \tag{26}$$

In (26), s_{ij} is similar to the membership function, representing the possibility that a pixel belongs to a category, and parameters p and q are the weight parameters that control the membership function and the spatial relationship function, respectively. Here, $u_{ij}^{'}$ is the final membership function. **Figure 5** demonstrates the flowchart of the improved fuzzy clustering algorithm SIPMFCM.

Furthermore, the brain image features are extracted. The U-Net algorithm only uses the features of the last convolutional layer, the information is simple, and it is not enough to contain rich detailed information. Hence, the following improvements are made. First, the introduction of the feature fusion allows the model to fuse information of multiple scales to assist in the image feature extraction. Second, fusing features of multiple scales helps to fuse semantic and location information, complements multiple feature information, and improves the accuracy of feature extraction. Consequently, HPU-Net can process multi-modal MR brain images. Besides, information of different scales is integrated to achieve effective and fast image segmentation.

Each Block in HPU-Net introduces BN (Batch Normalization). This neural network aims to learn the data distribution. Parameters are adopted to simulate data distribution to describe the original data and predict the unknown data (Wang J.L. et al., 2020). During training, the mean value μ_β of the internal data distribution of each batch is calculated first as Equation (27):

$$\mu_\beta = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{27}$$

In (27), m refers to the number of batches of the brain images. Then, the data distribution variance within each batch δ_β^2 is calculated as Equation (28):

$$\delta_\beta^2 = \frac{1}{m}\sum_{i=1}^{m}\left(x_i - \mu_\beta\right)^2 \tag{28}$$

Next, data in the current batch are normalized as Equation (29):

$$\hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\delta_\beta^2 + \varepsilon}} \tag{29}$$

BN can keep the feature with a mean value of 0 and a variance of 1, which is consistent with the input data distribution. Finally, BN also adds a step, "scale and shift," as described in Equation (30):

$$y_i = \gamma\hat{x}_i + \beta = BN_{\gamma,\beta}\left(x_i\right) \tag{30}$$

Equation (30) seems to be somewhat contrary to the previous operation; however, it can restore the original input. Besides, parameters γ,β shall meet Equation (31):

$$\begin{cases} \gamma^{(k)} = \sqrt{Var\left(x^{(k)}\right)} \\ \beta^{(k)} = E\left(x^{(k)}\right) \end{cases} \tag{31}$$

In this way, the entire model may or may not change the input distribution of the current layer in the end. The overall effect is to maintain the consistency of the feature distribution to stabilize the model during training and learning. Meanwhile, BN also helps accelerating the convergence time during model training, prevents gradient dispersion or gradient explosion, and improves training accuracy.

Simulation Experiment

MATLAB is adopted for simulation experiments to verify the performance of the constructed brain image processing and brain

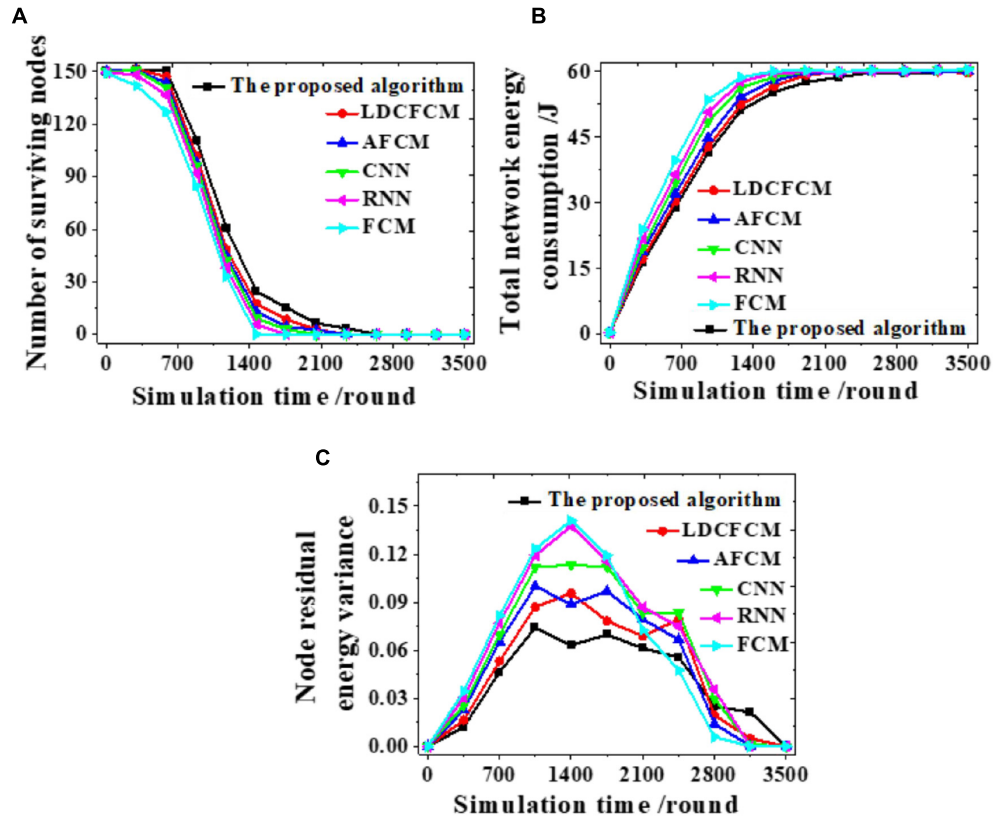


FIGURE 6 | The number of nodes surviving, the total energy consumption of the network, and the node remaining energy variance during the clustering process. **(A)** The number of nodes surviving. **(B)** The total energy consumption of the network. **(C)** The remaining energy variance of the nodes.

disease diagnosis prediction model based on improved fuzzy clustering and HPU-Net. The brain image data obtained in the experiment were all from a Hospital. Data of each patient is standardized to zero mean and unit standard deviation; then, slices that do not contain tumor information are deleted. The image size is cropped to 160 × 160 to reduce data imbalance. To validate the model's validity, data are divided into five parts equally; then, fivefold cross-validation is performed; that is, four sets of data are used for training and the remaining set of data is used for testing to obtain a set of experimental results. By analogy, different sets of data are adopted to make a test set separately so that five sets of experimental results can be obtained. The average of the five sets of experimental results is taken as the eventual result. The following hyperparameters shall be set for the neural network: 80 iterations, 3,500 rounds of simulations, and 128 batch size. In the performance analysis, the algorithm model proposed is compared with the algorithms applied by other scholars in related fields, CNN (Wu and Cao, 2021), RNN (Roy and Maji, 2020), FCM, LDCFCM (Local Density Clustering Fuzzy C-Means) (Dogra et al., 2020), and AFCM (Adaptive Fuzzy C-Means) (Song et al., 2018) algorithms are selected, respectively, for comparative analysis regarding node survival number, total energy consumption, and residual energy variance. In this simulation, the deep learning algorithm library Keras is utilized, which is an advanced neural network API written in Python with Tensorflow as the back end. Adam is utilized as the optimizer; the standard backpropagation is adopted to train the model, and all parameters are initialized randomly. The computer configuration is Ubuntu 16.04, the CPU (Central Processing Unit) is ES-2640 2,566 memory, and the GPU (Graphics Processing Unit) is GTX 1080Ti.

In brain image feature extraction, the similarity between the segmentation results and the standard expert segmentation results measures the quality of the segmentation results (Debakla et al., 2019). Segmentation evaluation criteria include DSC (Dice Similarity Coefficient), representing the degree of similarity between the segmentation result of the experiment and the expert's label, ranging from 0 to 1. The higher the DSC value is, the higher the segmentation similarity is. The Jaccard coefficient and DSC are defined as Equations (32) and (33):

$$DSC(T, P) = \frac{2\,|T \wedge P|}{|T| + |P|} \tag{32}$$

$$Jaccard(T, P) = \frac{|T \wedge P|}{|T \vee P|} \tag{33}$$

In (32) and (33), T represents the original tumor area, and P represents the segmentation result of the proposed algorithm. While evaluating the segmentation results of the current model, P and T need to be comprehensively considered.

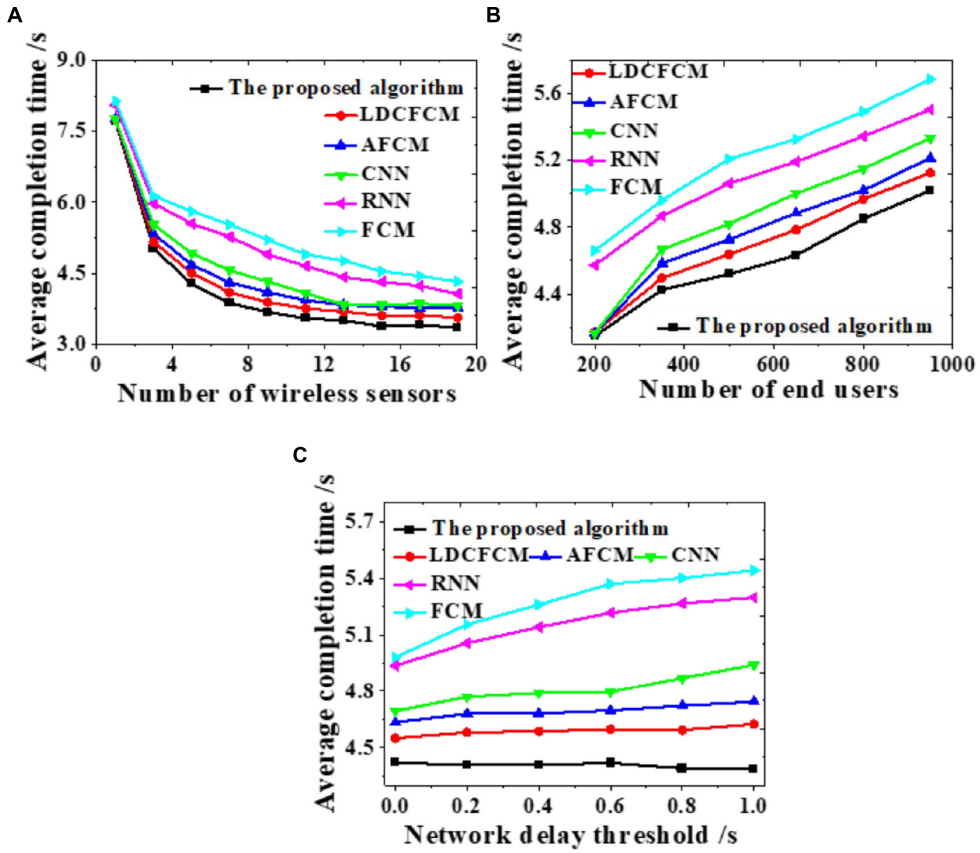


FIGURE 7 | Influences on the average completion time under different factors. **(A)** Number of wireless sensors. **(B)** Number of end-users. **(C)** Network delay threshold.

The evaluation coefficient V_{pc} and segmentation entropy V_{pe} are included for analysis and evaluation (Huang et al., 2020). The idea of this evaluation function is that if the fuzzy clustering effect is better, the membership difference of the same pixel belonging to different categories will be greater, the segmentation coefficient obtained will be larger, and the segmentation entropy will be smaller.

$$V_{pc} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^2}{n} \tag{34}$$

$$V_{pe} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{c} \left[u_{ij}^2 \log u_{ij} \right]}{n} \tag{35}$$

RESULTS AND DISCUSSION

Clustering Process Feasibility

Different models are analyzed from the number of nodes surviving, the total energy consumption of the network, and the node remaining energy variance to analyze the feasibility of the proposed algorithm. **Figure 6** presents the results.

The clustering process of each model is analyzed. The number of nodes surviving decreases as the number of simulation rounds increases. When the proposed algorithm runs to 2,637 rounds, the number of nodes survived is 0. In comparison, the number of nodes surviving of other models all reduce to 0 at an earlier time (**Figure 6A**). Changes in remaining energy consumption reveal that the energy consumption of each round of the network is reduced after the algorithm is used for clustering. When the total energy consumption of the network is exhausted, the system simulation reaches 2,560 rounds. In contrast, the total energy consumption of the network of other algorithms is faster, showing that fuzzy clustering reduces the energy consumption of the network (**Figure 6B**). Moreover, the proposed algorithm fluctuates within 0∼0.070, and the curve changes smoothly. Whereas, the fluctuation range of other models all exceeds 0.070 (**Figure 6C**). Therefore, the proposed algorithm can provide more nodes survive, lower energy consumption, and more stable changes under the same conditions.

Overall Network Performance

Influences of the number of wireless sensors, the number of end-users, and changes in network delay threshold on the average task completion time are evaluated to analyze the overall network performance, as displayed in **Figure 7**.

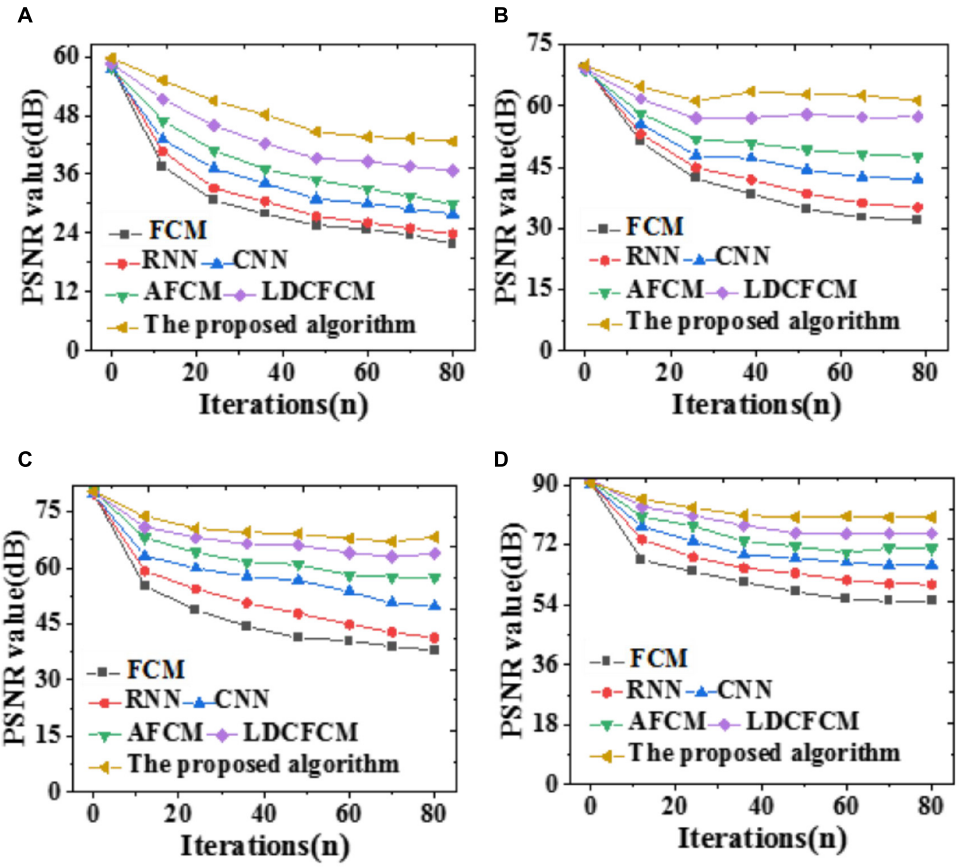


FIGURE 8 | PSNR of the model segmentation results under different brain image noises. **(A)** 10%; **(B)** 20%; **(C)** 30%; **(D)** 40%.

As the number of wireless sensors increases, the average completion time of the system gradually decreases until it stabilizes (**Figure 7A**). As the number of end-users increases, the average completion time of system data transmission shows an upward trend (**Figure 7B**). As the network delay threshold increases, many models do not present notable changes. The average completion time of the proposed algorithm stabilizes at about 4.4 s, while the completion time of RNN and FCM shows an uptrend (**Figure 7C**). Hence, the proposed algorithm provides better overall network performance than other models all the time, whose average completion time of system data transmission is the shortest.

TABLE 1 | Changes in evaluation functions of different models (%).

	v_{pc}	v_{pe}
The proposed algorithm	0.948	0.051
LDCFCM	0.923	0.069
AFCM	0.864	0. 073
CNN	0.847	0.098
RNN	0.803	0.114
FCM	0.795	0.152

TABLE 2 | Variation of each model with the Jaccard coefficient (%).

	The white matter	The gray matter
The proposed algorithm	0.854	0.798
LDCFCM	0.839	0.742
AFCM	0.821	0. 711
CNN	0.764	0.679
RNN	0.736	0.593
FCM	0.617	0.561

Prediction Performance

PSNR (Peak Signal-to-Noise Ratio) measures a model's noise immunity. The larger the PSNR is, the more apparent the denoising effect is, and vice versa. In the present work, all models are compared under 10, 20, 30, and 40% image noises, as in **Figure 8**.

As the noise value increases, the denoising effect becomes apparent, and the improvement over algorithm PSNR is greater. There are many noise points after segmentation by FCM, which confuses noise and non-noise, resulting in a low PSNR value. Nevertheless, other similar algorithms have a poor removal effect on noise points, and their PSNRs are lower than those of the proposed algorithm. Sometimes, noises in the image seem to be

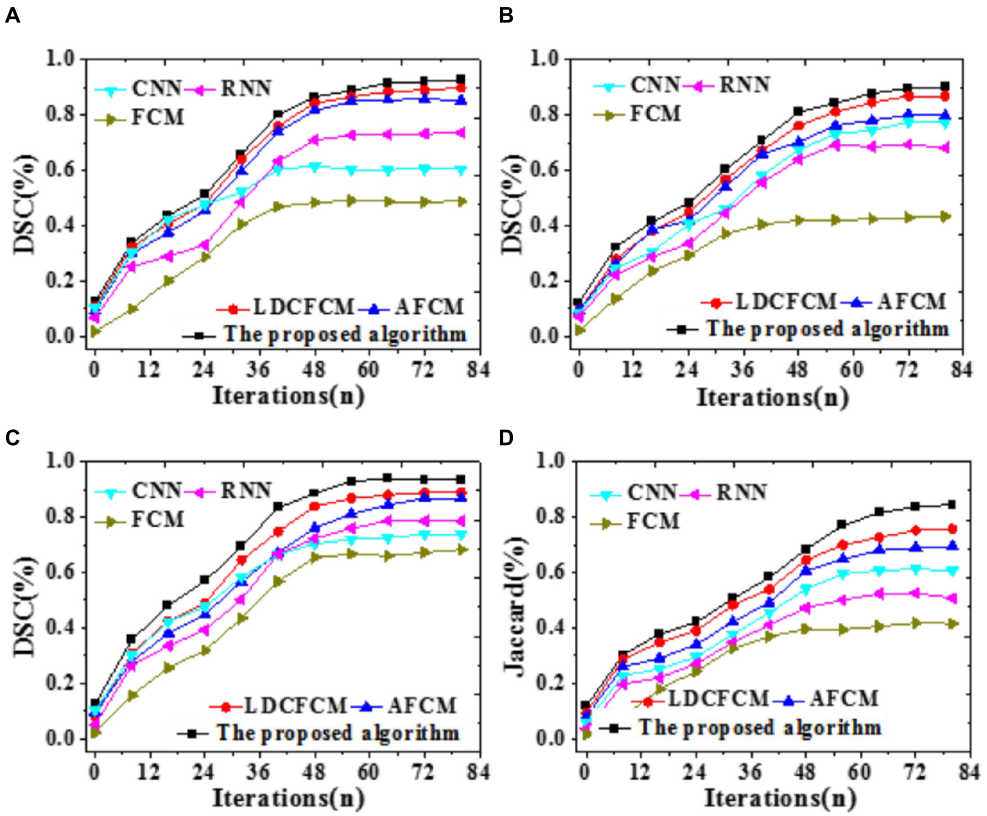


FIGURE 9 | Prediction accuracy indicators of different models for brain images with iterations. **(A)** Core tumor under DSC coefficient. **(B)** Enhancing tumor. **(C)** Whole tumor under DSC coefficient. **(D)** Jaccard coefficient.

suppressed, and PSNR seems to be improved; however, the edge and detail information of the image will be lost to varying degrees. Next, errors of each model are compared regarding evaluation functions and the Jaccard coefficient. The results are summarized in **Tables 1**, **2**.

Tables 1, **2** demonstrate that the proposed algorithm provides an evaluation coefficient V_{pc} of 0.948 and a segmentation entropy V_{pe} of 0.051, presenting a fuzzy clustering effect significantly better than other models. Regarding the Jaccard coefficient, the proposed algorithm provides recognition accuracies of 0.854 and 0.798 for the white matter and the gray matter on brain images, respectively, apparently better than other algorithms. Therefore, the proposed algorithm provides stronger recognition accuracy and robustness regarding feature recognition and diagnosis prediction.

To compare the prediction accuracy of different models, DSC is included to predict the recognition effect of the Whole Tumor, Core Tumor, and Enhancing Tumor of brain images; besides, the recognition effect of the Jaccard coefficient to brain images is analyzed, as illustrated in **Figure 9**.

Figure 9 reveals that DSC and Jaccard coefficient of the proposed algorithm are significantly improved. Under DSC, it provides the highest prediction accuracy for the Whole Tumor, reaching 0.936, followed by Core Tumor and Enhancing Tumor. The segmentation result of the Core Tumor only makes a small improvement under DSC. The Enhancing Tumor only contains the enhanced part of the tumor, occupying a small part of the tumor; sometimes, this area is segmented into other tumor parts. However, the size of the enhanced area is vital in clinical diagnosis. Moreover, the Jaccard coefficient of the proposed algorithm is 0.845, and the segmentation accuracy is remarkably better than that of other algorithms. Therefore, the proposed algorithm based on fuzzy clustering and HPU-Net outperforms other MRI brain tumor image segmentation proposed in recent years.

CONCLUSION

Brain tumor images have complicated edge structures and are prone to artifacts and offset fields, affecting image segmentation. The feature extraction and correct diagnosis of brain tumors in multi-sequence MRI images are particularly essential. The traditional fuzzy clustering algorithm is improved to solve the varying degrees of noises, weak boundaries, and artifacts in brain images. Moreover, a brain image processing and brain disease diagnosis prediction model is proposed based on HPU-Net and improved fuzzy clustering. Simulation experiments prove that the proposed algorithm can provide high accuracy for feature

extraction and recognition, excellent noise reduction effect, and the optimal image segmentation recognition effect. The highest prediction accuracy for DSC reaches 0.936, and that for the Jaccard coefficient reaches 0.845, providing an experimental basis for brain image feature recognition and brain disease prediction diagnosis. Still, there are some weaknesses in the present work. The proposed algorithm analyzes the overall characteristics while segmenting the brain tumor images, which may cause the difference between different individuals to be insignificant. In the future, texture features and shape features can be utilized for experiments; useful features can be left, and better segmentation results can be obtained by selecting multiple features. Eventually, the key steps to the precise treatment of brain tumors can be implemented to meet the actual needs of clinical medicine, which is vital for the subsequent clinical diagnosis and treatment of brain diseases.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

MH was mainly responsible for writing the main part of the full text. YZ was responsible for experimental operation and design of related experiments. SX was responsible for data analysis and data processing. HL was responsible for reviewing the full text, verifying the experimental results, and optimizing the structure of the article. ZL was responsible for the collection of relevant statistical data and the verification of the theoretical part. All authors contributed to the article and approved the submitted version.

REFERENCES

Ahmadi, N. (2020). A hybrid intelligent approach for image segmentation and feature extraction using fuzzy clustering, lattice boltzmann and GLDM techniques. *J. Soft Comput. Decis. Support Syst.* 7, 1–5. doi: 10.5815/ijigsp.2012.06.01

Bai, X., Zhang, Y., Liu, H., and Wang, Y. (2018). Intuitionistic center-free FCM clustering for MR brain image segmentation. *IEEE J. Biomed. Health Inform.* 23, 2039–2051. doi: 10.1109/jbhi.2018.2884208

Barricelli, B. R., Casiraghi, E., and Fogli, D. (2019). A survey on digital twin: definitions, characteristics, applications, and design implications. *IEEE Access* 7, 167653–167671. doi: 10.1109/access.2019.2953499

Bello, G. A., Dawes, T. J., Duan, J., Biffi, C., De Marvao, A., Howard, L. S., et al. (2019). Deep-learning cardiac motion analysis for human survival prediction. *Nat. Mach. Intell.* 1, 95–104. doi: 10.1038/s42256-019-0019-2

Debakla, M., Salem, M., Djemal, K., and Benmeriem, K. (2019). Fuzzy farthest point first method for MRI brain image clustering. *IET Image Process.* 13, 2395–2400. doi: 10.1049/iet-ipr.2018.6618

Dogra, J., Jain, S., Sharma, A., Kumar, R., and Sood, M. (2020). Brain tumor detection from MR images employing fuzzy graph cut technique. *Recent Adv. Comput. Sci. Commun.* 13, 362–369. doi: 10.2174/2213275912666181207152633

Du, Z., Kao, Y., and Park, J. H. (2019). Interval type-2 fuzzy sampled-data control of time-delay systems. *Inform. Sci.* 487, 193–207. doi: 10.1016/j.ins.2019.03.009

Feng, S., and Chen, C. P. (2018). Fuzzy broad learning system: a novel neuro-fuzzy model for regression and classification. *IEEE Trans. Cybern.* 50, 414–424. doi: 10.1109/tcyb.2018.2857815

Harifi, S., Khalilian, M., Mohammadzadeh, J., and Ebrahimnejad, S. (2020). Optimizing a neuro-fuzzy system based on nature-inspired emperor penguins colony optimization algorithm. *IEEE Trans. Fuzzy Syst.* 28, 1110–1124. doi: 10.1109/tfuzz.2020.2984201

Huang, Y. P., Singh, P., and Kuo, H. C. (2020). A hybrid fuzzy clustering approach for the recognition and visualization of MRI images of Parkinson's disease. *IEEE Access* 8, 25041–25051. doi: 10.1109/access.2020.2969806

Jeyaraj, P. R., and Nadar, E. R. S. (2019). Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J. Cancer Res. Clin. Oncol.* 145, 829–837. doi: 10.1007/s00432-018-02834-7

Jiang, Y., Zhao, K., Xia, K., Xue, J., Zhou, L., Ding, Y., et al. (2019). A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation. *J. Med. Syst.* 43, 1–9.

Kong, Y., Wu, J., Yang, G., Zuo, Y., Chen, Y., Shu, H., et al. (2019). Iterative spatial fuzzy clustering for 3D brain magnetic resonance image supervoxel segmentation. *J. Neurosci. Methods* 311, 17–27. doi: 10.1016/j.jneumeth.2018.10.007

Li, X., Cao, J., Liu, Z., and Luo, X. (2020). Sustainable business model based on digital twin platform network: the inspiration from haier's case study in China. *Sustainability* 12:936. doi: 10.3390/su12030936

Li, X., He, B., Wang, Z., Zhou, Y., Li, G., and Jiang, R. (2021). Semantic-enhanced digital twin system for robot–environment interaction monitoring. *IEEE Trans. Instrum. Meas.* 70, 1–13. doi: 10.1109/tim.2021.3066542

Liu, Y., Xu, Z., and Li, C. (2018). Distributed online semi-supervised support vector machine. *Inform. Sci.* 466, 236–257. doi: 10.1016/j.ins.2018.07.045

Mahata, N., and Sing, J. K. (2020). A novel fuzzy clustering algorithm by minimizing global and spatially constrained likelihood-based local entropies for noisy 3D brain MR image segmentation. *Appl. Soft Comput.* 90:106171. doi: 10.1016/j.asoc.2020.106171

Maqsood, A., Oslebo, D., Corzine, K., Parsa, L., and Ma, Y. (2020). STFT cluster analysis for DC pulsed load monitoring and fault detection on naval shipboard power systems. *IEEE Trans. Transp. Electrification* 6, 821–831. doi: 10.1109/tte.2020.2981880

Mishro, P. K., Agrawal, S., Panda, R., and Abraham, A. (2020). Novel fuzzy clustering-based bias field correction technique for brain magnetic resonance images. *IET Image Process.* 14, 1929–1936. doi: 10.1049/iet-ipr.2019.0942

Pham, T. X., Siarry, P., and Oulhadj, H. (2018). Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation. *Appl. Soft Comput.* 65, 230–242. doi: 10.1016/j.asoc.2018.01.003

Rahman, T., Ullah, I., Rehman, A. U., and Naqvi, R. A. (2020). Notice of violation of IEEE publication principles: clustering schemes in MANETs: performance evaluation, open challenges, and proposed solutions. *IEEE Access* 8, 25135–25158. doi: 10.1109/access.2020.2970481

Renard, F., Guedria, S., De Palma, N., and Vuillerme, N. (2020). Variability and reproducibility in deep learning for medical image segmentation. *Sci. Rep.* 10:13724.

Roy, S., and Maji, P. (2020). Medical image segmentation by partitioning spatially constrained fuzzy approximation spaces. *IEEE Trans. Fuzzy Syst.* 28, 965–977. doi: 10.1109/tfuzz.2020.2965896

Senthilkumar, C., and Gnanamurthy, R. K. (2019). A Fuzzy clustering based MRI brain image segmentation using back propagation neural networks. *Cluster Comput.* 22, 12305–12312. doi: 10.1007/s10586-017-1613-x

Song, J., Cong, W., and Li, J. (2018). A robust fuzzy c-means clustering model with spatial constraint for brain magnetic resonance image segmentation. *J. Med. Imaging Health Inform.* 8, 811–816. doi: 10.1166/jmihi.2018.2332

Wang, J. L., Qiu, S. H., Chen, W. Z., Wu, H. N., and Huang, T. (2020). Recent advances on dynamical behaviors of coupled neural networks with and without reaction–diffusion terms. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 5231–5244. doi: 10.1109/tnnls.2020.2964843

Wang, Z., Yao, L., Cai, Y., and Zhang, J. (2020). Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind

turbine rolling bearings fault diagnosis. *Renew. Energy* 155, 1312–1327. doi: 10.1016/j.renene.2020.04.041

Wu, C., and Cao, Z. (2021). Noise distance driven fuzzy clustering based on adaptive weighted local information and entropy-like divergence kernel for robust image segmentation. *Digital Signal Process.* 111:102963. doi: 10.1016/j.dsp.2021.102963

Xie, W., Liu, B., Bu, L., Wang, Y., and Zhang, J. (2020). A decoupling approach for observer-based controller design of TS fuzzy system with unknown premise variables. *IEEE Trans. Fuzzy Syst.* 1–1. doi: 10.1109/tfuzz.2021.3063818

Yoo, S. H., Geng, H., Chiu, T. L., Yu, S. K., Cho, D. C., Heo, J., et al. (2020). Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Front. Med.* 7:427. doi: 10.3389/fmed.2020.00427

Zhang, Y., Govindaraj, V., Murugan, P. R., and Sankaran, S. (2020). Smart identification of topographically variant anomalies in brain magnetic resonance imaging using a fish school based fuzzy clustering approach. *IEEE Trans. Fuzzy Syst.* 1–1. doi: 10.1109/tfuzz.2020.3015591

Zhao, F., Liu, H., Fan, J., Chen, C. W., Lan, R., and Li, N. (2018). Intuitionistic fuzzy set approach to multi-objective evolutionary clustering with multiple spatial information for image segmentation. *Neurocomputing* 312, 296–309. doi: 10.1016/j.neucom.2018.05.116

Zheng, Y., Wang, S., Li, Q., and Li, B. (2020). Fringe projection profilometry by conducting deep learning from its digital twin. *Opt. Express* 28, 36568–36583. doi: 10.1364/oe.410428

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

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

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

frontiers in Neuroscience

REVIEW
published: 19 August 2021
doi: 10.3389/fnins.2021.733546

A Review on Signal Processing Approaches to Reduce Calibration Time in EEG-Based Brain–Computer Interface

Xin Huang[1], Yilu Xu[2], Jing Hua[2], Wenlong Yi[2], Hua Yin[2], Ronghua Hu[3] and Shiyi Wang[4]*

[1] Software College, Jiangxi Normal University, Nanchang, China, [2] School of Software, Jiangxi Agricultural University, Nanchang, China, [3] School of Mechatronics Engineering, Nanchang University, Nanchang, China, [4] Youth League Committee, Jiangxi University of Traditional Chinese Medicine, Nanchang, China

In an electroencephalogram- (EEG-) based brain–computer interface (BCI), a subject can directly communicate with an electronic device using his EEG signals in a safe and convenient way. However, the sensitivity to noise/artifact and the non-stationarity of EEG signals result in high inter-subject/session variability. Therefore, each subject usually spends long and tedious calibration time in building a subject-specific classifier. To solve this problem, we review existing signal processing approaches, including transfer learning (TL), semi-supervised learning (SSL), and a combination of TL and SSL. Cross-subject TL can transfer amounts of labeled samples from different source subjects for the target subject. Moreover, Cross-session/task/device TL can reduce the calibration time of the subject for the target session, task, or device by importing the labeled samples from the source sessions, tasks, or devices. SSL simultaneously utilizes the labeled and unlabeled samples from the target subject. The combination of TL and SSL can take advantage of each other. For each kind of signal processing approaches, we introduce their concepts and representative methods. The experimental results show that TL, SSL, and their combination can obtain good classification performance by effectively utilizing the samples available. In the end, we draw a conclusion and point to research directions in the future.

Keywords: signal processing, transfer learning, semi-supervised learning, EEG, brain–computer interface, calibration

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Wenjuan Liao,
University of Colorado, United States
Zhen Shen,
Nanyang Institute of Technology, China

***Correspondence:**
Yilu Xu
xu_yilu@jxau.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *30 June 2021*
Accepted: *30 July 2021*
Published: *19 August 2021*

Citation:
Huang X, Xu Y, Hua J, Yi W, Yin H, Hu R and Wang S (2021) A Review on Signal Processing Approaches to Reduce Calibration Time in EEG-Based Brain–Computer Interface. Front. Neurosci. 15:733546. doi: 10.3389/fnins.2021.733546

INTRODUCTION

A brain–computer interface (BCI) can allow a subject to directly control an external electronic device using his brain signals without the participation of his peripheral nerves and muscles (Vidal, 1977; Kübler et al., 2001; Wolpaw et al., 2002). BCI technology can not only help the patients suffering from neuromuscular diseases, such as amyotrophic lateral sclerosis (ALS), recover or live independently, but also provide the healthy subjects with a novel way to communicate with the external environment (Chai et al., 2014; Horki et al., 2015). Therefore, BCI technology can play an important role in rehabilitation engineering, military, and entertainment (Ma et al., 2015; Soekadar et al., 2016).

So far, there are three categories of BCIs divided by the extent of invading the brain, including non-invasive BCIs, semi-invasive BCIs, and invasive BCIs (Rao, 2010). Due to safety, non-invasive BCIs have drawn increasing attention, which can collect non-invasive brain signals, such as electroencephalograms (EEGs), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS). Particularly, due to low cost and high temporal resolution, EEG signals are widely used in the non-invasive BCIs and much suitable for real-time BCI control.

However, EEG signals are weak, sensitive to noise/artifact, and non-stationary. It is challenging to classify EEG signals accurately. Moreover, it is difficult for a subject to freely control his brain activity to execute kinds of BCI tasks. On past decades, three classic EEG paradigms have been widely studied as follows:

(1) Motor imagery (MI) (Pfurtscheller and Neuper, 2001), which is a typical BCI paradigm, was firstly designed for rehabilitation treatment. The individuals with neuromuscular diseases are expected to recover their damaged motor nerves by effectively implementing MI tasks. Unfortunately, people are familiar with real movement naturally instructed by their brain activity, instead of movement imagination. Therefore, MI EEG signals are quite unsteady, limiting the categories of MI. The mostly used MI tasks include left hand, right hand, foot, and tongue MI (Kumar and Inbarani, 2017; She et al., 2020).
(2) Event-related potentials (ERP) (Blankertz et al., 2011; Zhang et al., 2012), which are invoked by an attended stimulus, have more obvious features than MI EEG signals. For example, the P300 ERP can be detected after a rare and relevant stimulus onset, with a latency around 300 ms. Farwell and Donchin (1988) demonstrated a P300 speller system with 36 commands and an information transfer rate (ITR) of less than 30 bits/min.
(3) Steady-state visual evoked potentials (SSVEP), which are evoked by a low-frequency flickering stimulus, oscillate at a multiple of the stimulus frequency (Yin et al., 2015; Adair et al., 2017; Chiang et al., 2019). SSVEP-based BCI has advantages, such as multiple commands, high classification performance, and high ITR. However, users may suffer from visual fatigue after long-term gazing at the flickering stimulus.

Besides these traditional paradigms, some meaningful paradigms become emerging in EEG-based BCI. For instance, emotion states (happy, neutral, and sad) can be detected in EEG-based affective BCI (Muhl et al., 2014; Shen and Lin, 2019). Drive drowsiness estimation from EEG signals can be used to avoid traffic accidents (Wei et al., 2015; Wu et al., 2016; Chai et al., 2017). Although EEG-based BCI has promising perspective, it must deal with the weakness of EEG and uncontrollability of BCI task. In this paper, we discuss the signal processing approaches to tackle this problem. A whole process of a traditional EEG-based BCI system is illustrated in **Figure 1**.

In **Figure 1**, the whole process is comprised of two consecutive phases: calibration phase and testing phase. In the calibration phase, labeled samples from each subject are successively processed by different modules to train a subject-specific classifier. Then, in the testing phase, an unlabeled sample is identified by the tailored classifier and translated into a command to control an external electronic device, e.g., a wheelchair or an exoskeleton arm. Here, we briefly introduce main modules in both phases as follows:

(1) Signal acquisition, which is used to acquire the EEG signals using multiple dry/wet electrodes and then digitize these electrical signals. It is a critical step for BCI systems. In the calibration phase, the raw EEG signals are invoked by assigned BCI tasks. A reliable classifier benefits from amounts of labeled samples. In the testing phase, the raw EEG signals are invoked by unknown human intentions and then identified in the classification module.
(2) Signal preprocessing, which aims to improve signal quality without losing important information. In this step, the raw EEG signals, easily contaminated by electrooculograms (EOGs), electromyograms (EMGs), and DC drift, are cleaned and denoized to enhance their relevant information. Principal component analysis (PCA) and independent component analysis (ICA) are often used to remove EMGs and EOGs from the raw EEG signals.
(3) Feature extraction, which aims at extracting a few relevant features from the preprocessed EEG signals to alleviate the computational burden of classification. The brain patterns can be characterized by these features. The preprocessed EEG signals are typically filtered in the time domain (band-pass filter), frequency domain, time-frequency domain, or spatial domain (spatial filter). The best filters optimized in the calibration phase are used to extract valuable features in the testing phase.
(4) Classification. In the calibration phase, a set of labeled features are input to the classification module to build a subject-specific classifier. In the testing phase, an unlabeled feature is assigned to a class by this tailored classifier. A class corresponds to one type of BCI tasks.

The neurophysiological processes often vary over time and across subjects in most EEG-based BCIs. For each subject, long calibration procedure is always needed to collect amounts of labeled EEG signals to build a steady recognition model. However, tedious calibration may cause user's mental exhaustion and degenerate the experimental effect. It is hard to get abundant EEG signals for each subject. In **Table 1**, some publicly available EEG-based datasets are listed.

As shown in **Table 1**, the number of total EEG trials is limited for each subject in most datasets. Accordingly, labeled samples are not enough to be exploited. Therefore, it is crucial to make full use of the samples available and face up to the differences among them. For each subject, the lack of labeled samples motivates the approaches that go beyond traditional supervised learning by importing labeled samples from other sessions/subjects/tasks/devices, unlabeled samples from himself, and even artificial samples (Lotte, 2015). These approaches include transfer learning (TL), deep learning (DL), semi-supervised learning (SSL), artificial data generation and so on.

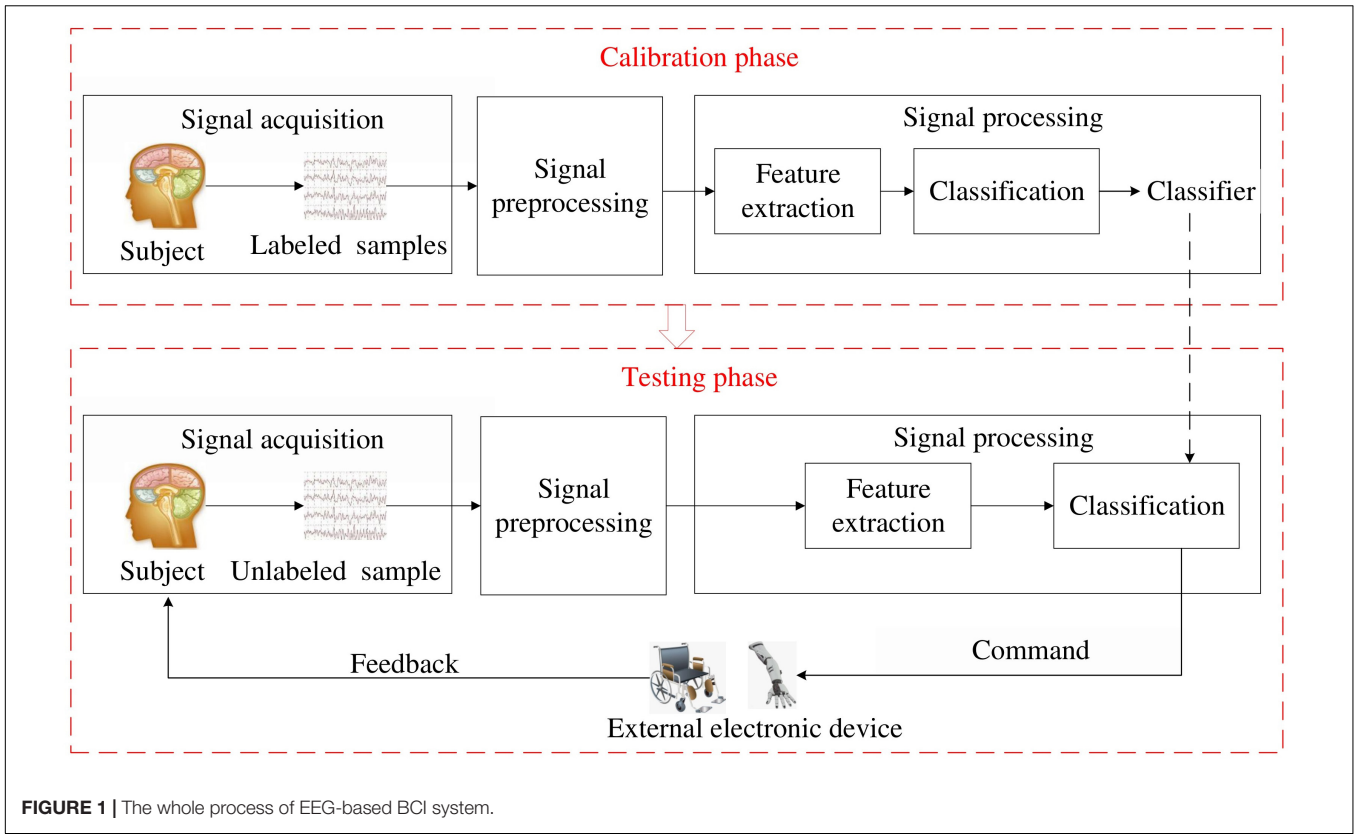


FIGURE 1 | The whole process of EEG-based BCI system.

TABLE 1 | Publicly available EEG-based datasets.

Dataset	Subject	Task	Total trials (per subject)	References
Dataset A/B	10/9	Left/right hand MI	300/90	Alamgir et al. (2010)
BCI III IVa	5	Right hand/foot MI	280	Blankertz et al. (2006)
BCI IV IIa	9	Left/right hand, foot, and tongue MI	288	Keng et al. (2012)
MI BCI	52	Left/right hand MI	200 or 240	Hohyun et al. (2017)
P300 speller	8	36 characters	35	Riccio et al. (2013)
SEED	15	Negative, neutral, and positive emotion states	15	Zheng and Lu (2015)

Transfer learning is a popular machine learning technique to shorten calibration time because it transfers abundant labeled samples from different sessions/subjects/tasks/devices, denoted as source domains, to a new session/subject/task/device, denoted as a target domain. The main hypothesis in TL is that the target and source domains belong to the same feature space. However, high inter-domain variability often makes the hypothesis violated. It is key to effectively integrate the features of the source domains with those of the target domain.

Deep learning is a promising subfield of machine learning, which has successful applications in many fields, such as natural language processing and computer vision, etc (Huang et al., 2021). Recently, it has attracted more attention in BCIs (Dose et al., 2018; Ming et al., 2018; Liu Y. et al., 2020; Zhang Y. et al., 2020). It can build a unified end-to-end model directly applied to raw EEG signals. However, its good performance strongly relies on sufficient labeled samples and a huge computational cost. Although DL cannot directly shorten the calibration time for a new BCI subject, its pre-trained model trained from a large scale of source domains can improve the performance of the BCI system. In the Section of TL, we discuss a combination of DL and TL.

Semi-supervised learning can simultaneously use the labeled and unlabeled samples from the same subject to train a classifier. For offline SSL, the inherent information embedded in the relatively large unlabeled set can effectively make up the insufficiency of the labeled set. For online SSL, the classifier can be updated in real time using a small, labeled set and an increasing unlabeled set with pseudo labels. To address with-in subject variability, SSL works by making the following assumptions, e.g., the smoothness assumption, the cluster assumption, and the low-density assumption. In general, with-in subject variability in SSL should be less than the inter-domain variability in TL. Therefore, it is valuable for SSL to reduce calibration time.

Recently, a combination of TL and SSL has been popular since it can take advantage of TL and SSL simultaneously. On the one

hand, it can make up the shortage of the labeled samples by using TL. On the other hand, it can train a more subject-specific model by importing the unlabeled samples from the target subject.

Artificial data generation is also a good choice to cope with the shortage of the labeled samples. Lotte (2015) generated numerous artificial labeled samples from the few original labeled samples available for EEG-based BCI by means of segmentation and recombination in the time domain or time-frequency domain. In the dataset 1 of BCI competition IV, a challenging problem was posed to discriminate between the artificial data and real data (Keng et al., 2012). The performance of artificial data generation strongly relies on the number and quality of original labeled data.

In this paper, we focus on commonly used approaches to shorten calibration effort, such as TL and SSL. Generally, TL and SSL are often used in feature extraction and classification, respectively. To our best knowledge, there is no comprehensive review on signal processing approaches to reduce calibration time in EEG-based BCI. Lotte (2015) performed a short review on the existing signal processing approaches to shorten or suppress calibration time and paid more attention to the introduction to their new methods. There are comprehensive reviews on the application of TL in EEG-based BCI (Wang et al., 2015; Zhang K. et al., 2020). However, few reviews on SSL are referred.

In this paper, we give an overview of signal approaches to minimize calibration time in EEG-based BCI. Meanwhile, the experimental results of representative approaches are analyzed. BCI illiteracy means that the classification accuracy of subject cannot reach 70%. To avoid BCI illiteracy, approximately 40 labeled samples per class are required for BCI subject, as suggested in Blankertz et al. (2007). Therefore, it is crucial for signal processing approaches to make full use of the data available to help the subject obtain satisfied performance using his labeled samples as few as possible.

The remainder of this paper is organized as follows. In Sections "Transfer Learning" and "Semi-supervised Learning," TL and SSL approaches are reviewed, respectively. In Section "A Combination of Transfer Learning and Semi-Supervised Learning," a combination of TL and SSL is introduced. A discussion of different signal processing approaches is presented in Section "Discussion." Finally, a conclusion is drawn in Section "Conclusion."

TRANSFER LEARNING

Concepts

Transfer learning is usually designed to cope with the shortage of the labeled data from the target domain by transferring the labeled data from other source domains. As mentioned above, a domain refers to a session, a subject, a task, or a device, etc. However, high between-domain variability in EEG-based BCI raises more problems: how to evaluate, minimize, and utilize such differences? To address these problems, we introduce the formal definition of TL and discuss state-of-the-art TL methods.

Definition 1. As in Pan and Yang (2010), a domain $\mathcal{D}$ consists of a feature space $\mathcal{X}$ and its marginal probability distribution $P(X)$, where $X \in \mathcal{X}$. A task $\mathcal{T}$ is defined as a label space $\mathcal{Y}$ and a predictive function $f(X)$. Although a source domain $\mathcal{D}_s$ do not match a target domain $\mathcal{D}_T$, or a source task $\mathcal{T}_s$ is different from a target task $\mathcal{T}_T$, i.e., $\mathcal{D}_s \neq \mathcal{D}_T$, or $\mathcal{T}_s \neq \mathcal{T}_T$, TL still aims to improve the learning level of the target predictive function $f_T(X)$ by adding the knowledge embedded in N_s source domains $\{\mathcal{D}_s^i\}_{i=1}^{N_s}$ and their associated tasks $\{\mathcal{T}_s^i\}_{i=1}^{N_s}$.

In this paper, state-of-the-art TL approaches in EEG-based BCI are categorized into three groups based on "What knowledge should be transferred" as follows (Pan and Yang, 2010):

(1) Instance transfer learning (ITL), which directly transfers the certain parts of the data (instance) in the source domains by reweighting. It is crucial to evaluate the similarity between the source and target domains and convert the corresponding similarity metric into optimal weights used in the source and target domains.
(2) Feature-representation transfer (FRT), which transfers the stationary feature representation from the source domains and adjusts the discriminative feature representation for the target domain. Unlike ITL, FRT learns "good" feature representations from the data in the source domains.
(3) Parameter transfer learning (PTL), which transfers the parameters of the model learning from the source domains based on the assumption that the source and target domains share these parameters. Then, PTL updates the parameters for the target domain.

In the following, we introduce some typical TL approaches belonging to different groups for analysis.

Instance Transfer Learning

In EEG-based BCI, ITL is widely used since it is easy to implement. In the phase of feature extraction, the samples from the source and target domains are usually reweighted based on the filters. Additionally, data alignment schemes have promising perspectives to minimize the differences between domains. Therefore, we introduce the filter-based ITL and data alignment-based ITL, respectively.

Filter-Based Instance Transfer Learning

Kullback-Leibler (KL) divergence is always used to compute similarities between two sets of EEG data. In weighted logistic regression-based TL (wLRTL) algorithm, the common spatial patterns (CSP) filtered EEG samples from each source subject were weighted based on KL divergence between the source and target subjects in a supervised way or an unsupervised way (Azab et al., 2019). In the supervised case, the divergence of each source subject is equal to the average of the KL divergences computed for each class. In the unsupervised case, the divergence of each source subject is calculated without using the class labels. Then, the regularization parameter calculated using the divergences of all source subjects is added to the objective function of logistic regression classifier to make full use of the labeled samples from the source subjects correctly. Cao et al. (2021) verified the applicability of wLRTL approach for stroke patients and intra/inter- subjective conditions.

Bhattacharyya distance was used to measure similarities between the source and target datasets based on their class conditional distributions in a probabilistic TL approach (Khalaf and Akcakaya, 2020). First, the CSP features of MI EEG trials from the source datasets are converted into scalar support vector machine (SVM) scores to obtain their class conditional distributions. Then, the similarity metrics of different source datasets are identified based on these class conditional distributions using Bhattacharyya distance. Finally, the top similar source datasets are combined with the target dataset to train the classifier.

Regularized CSP (RCSP) approaches have been very promising for cross-subject TL in the small-sample setting (Wang and Li, 2016). Based on the framework of RCSP, weighted covariance matrices of EEG data from the source and target subjects are simultaneously utilized to obtain optimal CSP filters. The distance metrics, including KL divergence, Frobenius norm, cosine distance, are used to measure the similarity between the source and target data (Kang et al., 2009; Cheng et al., 2017; Xu et al., 2019b).

A RCSP based on dynamic time warping (DTW-RCSP) approach was proposed to shorten temporal variations and non-stationarities between the subjects by aligning the labeled samples from all source subjects to the average of few target samples from the same class using an optimal warping path (Azab et al., 2020). This warping path is computed using dynamic programming. Then, using different weights, the subject-specific covariance matrix estimated using the labeled samples from the target subject is combined with the DTW-based transferred covariance matrix estimated using the aligned labeled samples from all source subjects mentioned above to get best CSP filters for the target subject.

The objective functions are often optimized to gain a weighted/projecting matrix/vector which can minimize the differences among domains (Dai et al., 2018; Zou et al., 2018; Jiao et al., 2019).

Zou et al. (2018) proposed two data mapping methods to diminish the inter-subject variation as much as possible.

In the first data mapping method, the transformation matrix $\mathcal{L}$ is optimized to reduce the inter-subject variation caused by dissimilarities in channel locations as follows:

$$\underset{\mathcal{L}}{\operatorname{argmin}} \left(\| W_1 - W_2\mathcal{L}^T \|_F^2 + \| \mathcal{L} - P \|_F^2 \right), \ st.\ \mathcal{L}^T\mathcal{L} = I \quad (1)$$

where W_1 and W_2 are the CSP filters calculated from the target domain D_1 and the source domain D_2, respectively. A new form D_2' which has similar channel locations with D_1 is generated by projecting $\mathcal{L}$ to original source domain D_2, i.e., $D_2' = \mathcal{L}D_2$. P is the constraint matrix with the element $P_{ij} = e^{-d(i,j)}$, where $d(i, j)$ is the distance between the channel i and the channel j. I is the identity matrix.

In the second data mapping method, the inter-subject variation of intensity of event-related desynchronization/synchronization (ERD/ERS) is minimized. CSP is designed to find the variance of the EEG signals between different subjects, instead of different classes. Therefore, the eigenvector of the CSP filter matrix is used to generate a correction matrix which can adjust the source data based on the target data.

Jiao et al. (2019) proposed a sparse group representation model (SGRM), which tried to seek the optimal representation vector for an unlabeled testing sample from the target subject by effectively using the labeled training samples from the target and source subjects. Each subject is regarded as a group. SGRM can realize not only the group-wise sparsity with the $L_{2,1}$-norm but also within-group sparsity with the L_1-norm. The representation vector u^* is optimized as:

$$\underset{u^*}{\operatorname{argmin}} \left(\frac{1}{2} \| \tilde{G}u^* - \hat{g} \|_2^2 + \lambda_1 \| u^* \|_1 + \lambda_2 (\| u_0 \|_2 + \sum_{i=1}^{N_s} \| u_i \|_2) \right), \quad (2)$$

where $\hat{g}$ denotes the CSP feature of the testing sample, $\tilde{G}$ concatenates the dictionary matrices from all subjects. The regularization terms λ_1 and λ_2 aim to control the within-group sparsity and the between-group sparsity, respectively. N_s denotes the number of source subjects. u_0 and u_i correspond to the sparse representation vectors of the target subject and the i-th source subject, respectively.

Data Alignment-Based Instance Transfer Learning

Recently, Riemannian alignment-based ITL approaches have drawn increasing attention in EEG-based BCI (Barachant et al., 2013; Qi et al., 2018; Zanini et al., 2018; Singh et al., 2019; Zhang and Wu, 2020). The reason is that the affine transformation makes data from different domains similar.

Covariance matrix is symmetric and positive semi-definite. It is positive definite when the sampling rate is high enough (Barachant et al., 2013). It is shown that symmetric positive definite (SPD) matrices belong to the smooth Riemannian manifold (Moakher, 2006). Therefore, as EEG features, covariance matrices are often used in the Riemannian manifold.

In **Figure 2**, the covariance matrices C_1 and C_2 are the two points of the Riemannian manifold. According to the congruence invariance property, the Riemannian distance $\delta(C_1, C_2)$ between C_1 and C_2 is invariant with respect to a change of reference matrix M. In general, M is the Riemannian mean of a specific domain, e.g., a class, a session, or a subject, etc.

After choosing M as the reference matrix, the original covariance matrices $\{C_i^\kappa\}_{i=1}^{N_\kappa}$ from the κ-th domain with N_κ trials are converted into aligned ones, i.e., $\{\tilde{C}_i^\kappa = M^{-1/2} C_i^\kappa M^{-1/2}\}_{i=1}^{N_\kappa}$ which center on the identity matrix. Each aligned covariance matrix is approximately an identity matrix, which can shorten the differences of each other.

A Riemannian geometry cross-session/subject TL framework was built in MI-based and P300-based BCIs (Zanini et al., 2018).

Firstly, some resting-state signals in each domain are collected to calculate their Riemannian mean. The resulting Riemannian mean is assigned as the reference matrix of corresponding domain. In MI-based BCI, resting signals are extracted from time windows of 1.5 s, where no BCI tasks are performed. However, in P300-based BCI, since there is not a separated

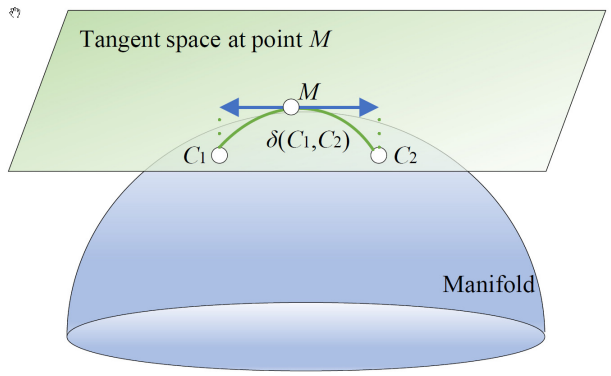


FIGURE 2 | A Riemannian manifold and its tangent space.

resting-state signal, reference matrix is calculated using the non-target trials.

Then, after Riemannian alignment (RA), the original covariance matrices from each domain are converted into aligned ones using the reference matrix of corresponding domain. All aligned labeled covariance matrices from source domains are used to estimate the centers of different classes and train a minimum distance to mean (MDM) classifier based on Riemannian Gaussian distributions.

Finally, the aligned testing covariance matrix from the target domain can be assigned to a class by the proposed classifier, considering its Riemannian distance to the center of class and the maximum likelihood estimator of the dispersion parameter of the class simultaneously.

Zhang and Wu (2020) proposed a cross-subject manifold embedded knowledge transfer (MEKT) approach. It is challenging and meaningful for MEKT to transfer the labeled source domains into the unlabeled target domain, which can boost zero-training for a new subject. Three steps of MEKT are briefly introduced as follows:

(1) Covariance matrix centroid alignment: MEKT separately performs RA for the covariance matrices in the source and target domains using their Riemannian means as corresponding reference matrices, so that the marginal probability distributions from different domains are similar.
(2) Tangent space (TS) feature extraction: MEKT converts the aligned covariance matrices from different domains into the tangent space feature vectors. Then, MEKT assembles the tangent feature vectors from the source domains.
(3) Mapping matrices identification: MEKT finds optimal projection matrices A for the labeled source feature vectors X_s and B for the unlabeled target feature vectors X_T to minimize the joint probability distribution shift between domains.

Finally, a shrinkage linear discriminant analysis (shrinkage LDA, sLDA) classifier is trained using the labeled source feature vectors $A^T X_s$ and applied to the unlabeled target feature vectors $B^T X_T$ to predict their classes. To further avoid negative transfer,

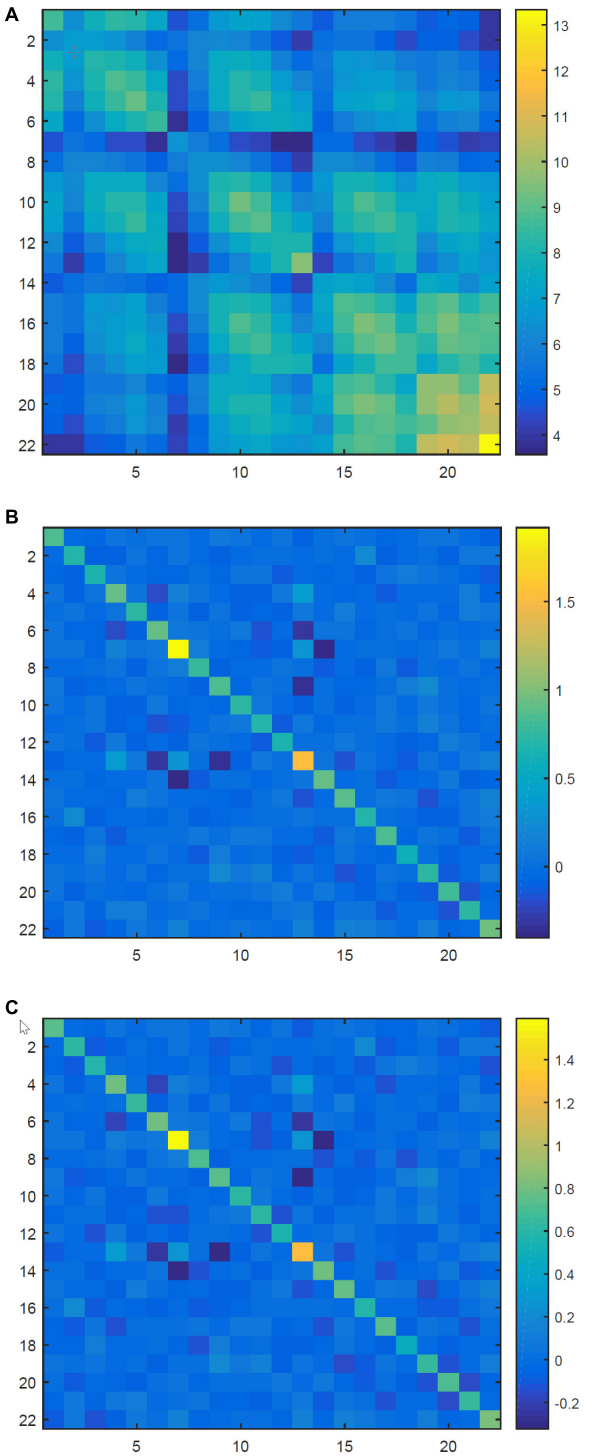


FIGURE 3 | The raw covariance matrix and different aligned matrices. **(A)** The raw covariance matrix. **(B)** The aligned covariance matrix after RA. **(C)** The aligned covariance matrix after EA.

Zhang and Wu (2020) put forward a domain transferability estimation (DTE) approach based on MEKT, which selected most similar source subjects before performing the second step above.

The transferability of the i-th source subject $\mathcal{D}_s^i$ is defined as follows:

$$transferability\left(\mathcal{D}_s^i, \mathcal{D}_T\right) = \frac{\| S_b^{\mathcal{D}_s^i} \|_1}{\| S_b^{\mathcal{D}_s^i, \mathcal{D}_T} \|_1}, \quad (3)$$

where $S_b^{\mathcal{D}_s^i}$ denotes the between-class scatter matrix of $\mathcal{D}_s^i$, and $S_b^{\mathcal{D}_s^i, \mathcal{D}_T}$ denotes the scatter matrix between the i-th source subject and the target subject $\mathcal{D}_T$. Then, the source subject with the highest transferability is chosen to be transferred.

Likewise, to select suitable source subjects, after RA, the aligned covariance matrices from a target subject are temporally used as the testing data and classified by an MDM classifier, which is trained using the aligned covariance matrices from a candidate source subject (Yong and Yu, 2020). The source subjects with higher classification performance are selected.

A Riemannian Procrustes analysis (RPA) approach was proposed using three geometrical transformations (translation, scaling, and rotation) in sequence (Rodrigues et al., 2019). During transformations, the covariance matrices from different domains have went through the steps of re-centering, stretching, and rotation. RPA is a typical semi-supervised TL method, utilizing the labeled data C_{sl} from the source domain, the labeled data C_{Tl} and the unlabeled data C_{Tu} from the target domain altogether.

In the translation phase, as in Zanini et al. (2018), the source labeled data C_{sl} and the target data $C_T = C_{Tl} \cup C_{Tu}$ are separately re-centered around the identity matrix using the resting activity of each session to compute the Riemannian means of different domains.

Then, in the scaling phase, the re-centered source labeled data $C_{sl}^{(rct)}$ and target labeled data $C_{Tl}^{(rct)}$ are used to calculate the ratio of their dispersions, namely the scaling factor r. The stretched source labeled data $C_{sl}^{(str)}$ and target data are $C_T^{(str)}$ are formed using the above scaling factor r, i.e., $C_{sl}^{(str)} = (C_{sl}^{(rct)})^r$, $C_T^{(str)} = (C_T^{(rct)})^r$, where $C_T^{(rct)} = C_{Tl}^{(rct)} \cup C_{Tu}^{(rct)}$. It should be noted that the re-centering and stretching transformations, without using the trials' labels, work in an unsupervised way.

Finally, in the rotation phase, the trials' labels are needed to compute an orthogonal rotation matrix U, which can be optimized to minimize the distance between the class mean of the source domain and that of the target domain. The rotated datasets are transformed as follows:

$$C_{Sl}^{(rot)} = U^T C_{Sl}^{(str)} U, C_T^{(rot)} = U^T C_T^{(str)} U. \quad (4)$$

After three geometrical transformations for the covariance matrices from different domains, the MDM classifier is used to infer the label of the unlabeled data from the target domain.

Since most classifiers are designed for the Euclidean space, instead of the Riemannian space, Euclidean alignment (EA) extends RA in the Euclidean space using the Euclidean mean as the reference matrix (He and Wu, 2020b). For each subject, the original covariance matrices are transformed into the aligned ones using his resting activities to calculate his Euclidean mean. The concatenated aligned training covariance matrices from different source subjects are filtered by CSP and then trained by LDA. The CSP features of the target subject are extracted from their aligned covariance matrices by CSP and assigned to a class by LDA.

Mathematically, the Euclidean mean is the arithmetic mean of all covariance matrices. And the Riemannian mean can be computed by iterating the following three steps until convergence: converting all covariance matrices in the manifold into the tangent space feature vectors, calculating the arithmetic mean of such feature vectors, and converting the arithmetic mean back into the covariance matrix in the manifold. Apparently, EA is simpler and faster than RA.

Furthermore, both RA and EA can make the aligned covariance matrices comparable since they are close to the identity matrix. To testify this, in **Figure 3**, we depict a raw covariance matrix and different aligned covariance matrices for the first EEG trial of A07 in dataset BCI IV IIa (Keng et al., 2012). As in Zhang and Wu (2020), this dataset contained 22-channel EEG recordings from nine healthy subjects. All raw EEG signals were band-pass filtered between 8 and 30 Hz using a causal 50-order finite impulse response filter. Then, the filtered signals were extracted from time segments ranging from 0.5 to 3.5 s. For each subject, their raw covariance matrices can be transformed into the aligned ones using Riemannian mean or Euclidean mean of their corresponding session as the reference matrix.

In **Figure 3A**, the raw covariance matrix is not like the identity matrix at all. However, as shown in **Figures 3B,C**, the aligned matrices are approximately identity matrix after RA and EA, respectively.

A label alignment (LA) was proposed to tackle cross-task transfer when the source and target subjects had different label spaces (He and Wu, 2020a). It is assumed that different subjects have the same number of classes. LA first performs EA for the per-class covariance matrices of each source subject. Then, LA builds k-medoids clustering on the unlabeled target domain based on a few labeled target trials and Riemannian distance. Then, LA labels the k cluster centers and performs EA for the per-class covariance matrices of the target subject. Next, LA matches each label in the source domain with one label in the target domain regularly or randomly, which depends on the similarity between label spaces. Finally, LA seeks a transformation matrix for each class of the source subject which can reduce the differences between the source subject's covariance matrices belonging to the specific class and the target subject's covariance matrices belonging to the matched class.

Feature-Representation Transfer

The main idea of FRT is to transfer the stationary features across different source domains and adjust the discriminative features for the target domain (Sybeldon et al., 2017; Yin et al., 2017; Hossain et al., 2018; Jin et al., 2018).

So far, there are many TL approaches using spatial filters to generate feature representations for the raw EEG data (Nakanishi et al., 2019; Zheng X. et al., 2020; Hehenberger et al., 2021). Nakanishi et al. (2019) proposed a task-related component analysis (TRCA) approach in a cross-device SSVEP-based BCI speller system. TRCA consists of two steps: extracting task-related

spatial filters for a source device and extracting task-unrelated spatial filters for a target device.

Step 1:

Task-related spatial filters are designed to maximize their reproducibility during the task periods. For each visual stimulus in SSVEP-based BCI, a specific task-related filter $\grave{w}$ is optimized by maximizing the inter-trial correlation of its projections $z(t) = \grave{w}^t x(t)$, where $x(t) \in \mathbb{R}^{N_{c_1}}$ denotes the t-th sample point of an observed EEG trial x collected from N_{c_1} channels in a source device. Two devices have inconsistent electrode montages.

The covariance matrix C_{h_1,h_2} between the projection $z^{(h_1)}$ of the h_1-th trial $x^{(h_1)}$ and the projection $z^{(h_2)}$ of the h_2-th trial $x^{(h_2)}$ is defined as follows:

$$C_{h_1,h_2} = Cov(z^{(h_1)}, z^{(h_2)}) = \sum_{k_1,k_2=1}^{N_{c_1}} \grave{\omega}_{k_1} \grave{\omega}_{k_2} Cov(x_{k_1}^{(h_1)}, x_{k_2}^{(h_2)}), \tag{5}$$

where $\grave{\omega}_{k_1}, \grave{\omega}_{k_2} \in \grave{w}$, $z^{(h_1)}$, $z^{(h_2)} \in \mathbb{R}^{N_{sa}}$ and $x^{(h_1)}$, $x^{(h_2)} \in \mathbb{R}^{N_{c_1} \times N_{sa}}$, N_{sa} denotes the number of samples in each trial. Then, for the specific visual stimulus, all covariance matrices between different labeled training trials are summed as:

$$\begin{aligned} \sum_{h_1,h_2=1,h_1\neq h_2}^{N_{tr}} C_{h_1,h_2} &= \sum_{h_1,h_2=1,h_1\neq h_2}^{N_{tr}} \sum_{k_1,k_2=1}^{N_{c_1}} \grave{\omega}_{k_1} \grave{\omega}_{k_2} Cov(x_{k_1}^{(h_1)}, x_{k_2}^{(h_2)}) \\ &= \grave{w}^T S \grave{w}, \end{aligned} \tag{6}$$

where N_{tr} is the number of labeled training trials, the element $S_{k_1,k_2}(1 \leq k_1, k_2 \leq N_{c_1})$ of S is formulated as:

$$S_{k_1,k_2} = \sum_{h_1,h_2=1,h_1\neq h_2}^{N_{tr}} Cov(x_{k_1}^{(h_1)}, x_{k_2}^{(h_2)}). \tag{7}$$

The optimization problem is defined as:

$$\widehat{\grave{w}} = \underset{\grave{w}}{\operatorname{argmax}} \frac{\grave{w}^T S \grave{w}}{\grave{w}^T Q \grave{w}}, \tag{8}$$

under the following constraint:

$$Var(z(t)) = \sum_{k_1,k_2=1}^{N_{c_1}} \grave{\omega}_{k_1} \grave{\omega}_{k_2} Cov(x_{k_1}(t), x_{k_2}(t)) = \grave{w}^T Q \grave{w} = 1, \tag{9}$$

where Var $(z(t))$ is the variance of $z(t)$. Then, for the n-th visual stimulus, its optimal task-related filter $\grave{w}_n$ is the eigenvector corresponding to the largest eigenvalue of $\widehat{\grave{w}}$ and its averaged training trials are $\overline{X}_n$. The N_f task-related filters are applied to their corresponding averaged training trials, i.e., $z = \grave{w}^T \overline{X}$, where $\grave{w} = [\grave{w}_1, \grave{w}_2, \cdots \grave{w}_{N_f}]$ and $\overline{X} = [\overline{X}_1, \overline{X}_2, \cdots \overline{X}_{N_f}]$. N_f denotes the number of visual stimuli.

Step 2:

Task-unrelated spatial filters $\hat{w} \in \mathbb{R}^{N_{c_2} \times N_f}$ used for a target device can be estimated as $\hat{w} = (xx^T)^{-1} xz^T$, where N_{c_2} and $x \in \mathbb{R}^{N_{c_2} \times N_{sa}}$ are the number of channels and a single trial in the target device, respectively. Likewise, task-unrelated spatial filters $\hat{w}$ are applied to the single trial x, i.e., $\hat{z} = \hat{w}^T x$. Obviously, z and $\hat{z}$ are the stationary features and the discriminative feature, respectively.

Zheng X. et al. (2020) proposed a cross-task TL approach in MI-based system, which can add more MI tasks based on the traditional ones. For example, the traditional single-limb task, such as left hand (LH) MI, right hand (RH) MI, or foot (F) MI, can be extended to be multiple-limb task, such as both left hand and right hand (LH and RH) MI simultaneously, both left/right hand MI and feet (LH/RH and F) MI simultaneously, etc. To tackle cross-task TL, stationary spatial filters can be first computed for the source domain using CSP. Then, the fishier ratio is used to select discriminative spatial filters for the target domain from stationary spatial filters.

Parameter Transfer Learning

The goal of PTL is to discover model's shared parameters from previous source domains (Sannelli et al., 2016).

Zheng M. et al. (2020) proposed a cross-session/subject model using the shared parameters and the specific parameter jointly. First, the source datasets which are most relevant to a new target dataset are selected based on Euclidean distance. Then, the shared parameters $\acute{W} = [\acute{\omega}_1, \acute{\omega}_2, \cdots, \acute{\omega}_s]$ are optimized using s selected source datasets as follows:

$$\begin{aligned} \underset{\acute{W},\acute{\mu},\acute{\Sigma}}{\operatorname{argmin}} \frac{1}{\lambda} \sum_s \| X_s \acute{\omega}_s - Y_s \|^2 + \sum_s \frac{1}{2} [(\acute{\omega}_s - \acute{\mu})^T \Sigma^{-1} (\acute{\omega}_s - \acute{\mu}) \\ + logdet(\acute{\Sigma})], \end{aligned} \tag{10}$$

where $\acute{\omega}_s$ denotes the model parameter for the specific source dataset (X_s, Y_s), $\acute{\mu}$ and $\acute{\Sigma}$ are, respectively, the mean and variance of all model parameters, λ refers to a regularization term. After iteratively updating $\acute{\mu}$ and $\acute{\Sigma}$, the shared parameters $\acute{W}$ converges. Finally, the shared parameters and the specific parameter trained in the target dataset are jointly utilized.

Domain adaption is also commonly used in PTL. The model parameter trained in the source domain can be reused and adjusted for the target domain (Wang et al., 2015).

Dai et al. (2019) proposed a domain transfer multiple kernels boosting (DTMKB) framework by applying the boosting techniques for learning multiple kernel-based classifiers. First, these kernel-based classifiers have already trained with the labeled samples from different source subjects. Then, a series of boosting labeled trials from the target subject are input to those classifiers one by one. For each boosting labeled trial, the best kernel classifier is selected and weighted based on the error rate. Finally, all best classifiers and their weights are combined to become an ensemble classifier.

Wei et al. (2018) proposed a cross-subject framework for drowsiness estimation by ranking and fusing different source models. First, for each source subject, an EEG-based drowsiness estimation source model is constructed using PCA and regression. Then, to rank different source models, the following steps are performed:

(1) To estimate the similarity among the source subjects, five distance metrics, including Euclidean distance, correlation

distance, Chebyshev distance, cosine distance, and KL divergence, are involved to calculate the distance between subjects' alert baseline power distributions and construct a multiple distance measurement matrix M.

(2) To further evaluate the transferability among source subjects, a transferability matrix XP can be obtained by calculating the performance of one source model on another source subject. Both $\mathcal{M}$ and XP are normalized into z-scores vector by vector.

(3) A linear support vector regression (SVR) model is trained using $\mathcal{M}$ and XP as the predictor and the response, respectively. For a new target session, the transferability of each source model is estimated based on the trained SVR model using the baseline distances between the corresponding source subject and the target subject. Then, the different source models are ranked from high to low based on the transferability.

Finally, the rankings of all source models are used to generate their weights in a fused model.

Jin et al. (2020) proposed a cross-subject generic model set in P300-based BCI. The P300 signals from 116 source subjects are used to train this generic model set without the participation of the target subject. The initial feature vectors are extracted from these source subjects' EEG signals. Then, the dimensionality of these feature vectors is reduced using PCA to alleviate computational burden. The resulting feature vectors are grouped into ten clusters using a K-means clustering algorithm. Ten weighted LDA (WLDA) classifiers are separately built using the feature vectors from ten clusters. In each WLDA classifier, its discriminant vector is initially trained by traditional LDA classifier using the feature vectors from each cluster. The vectors are then weighted based on their classification accuracies computed in the initial WLDA classifier. Next, the discriminant vector is updated by LDA using the weighted vectors. As a result, a generic model set consists of these ten WLDA classifiers.

Currently, a combination of TL and DL has gained attention in the BCI systems (Fahimi et al., 2019; Bird et al., 2020). TL focuses on the transfer of knowledge between different domains. DL aims to find the inherent knowledge among amounts of labeled trials. If the pre-trained model trained by DL is reliable and adjustable, it is meaningful to transfer the model from one domain to the other one.

Fahimi et al. (2019) developed a cross-subject TL approach based on an end-to-end deep convolutional neural network (CNN) to decode the attentive mental state for EEG-based BCI. The proposed method learns from raw EEG signals to avoid the loss of information and builds a network with the combination of convolutional, max-pooling, and dropout layers using a large scale of EEG data from 120 healthy source subjects. To ensure positive transfer, half of the target subject's samples are re-trained for adaption by CNN.

The summary of some representative TL approaches is shown in **Table 2**. In the sixth column of **Table 2**, 'Training' denotes the size of all labeled training sets. The first term of 'Training' refers to the size of labeled training set from the target domain, the second term of 'Training' refers to the product of the size of labeled training set from one source domain and the number of source domains, whereas 'Testing' denotes the size of unlabeled testing set from the target domain. CA denotes the classification accuracy. BCA denotes the balanced classification accuracy. Different from CA, BCA first calculates the classification accuracy for each class and then averages all classification accuracies. CC denotes the correlation coefficient between actual and predicted drowsiness indices.

As shown in **Table 2**, MI-based BCIs have been widely studied in the field of TL. The reason is that there are a lot of publicly available MI datasets. Furthermore, MI EEG signals are evoked by spontaneous movement imagination without external stimulus. Therefore, compared with ERP and SSVEP, MI is more uncontrollable and more eager to minimize the calibration time by using TL methods. In addition, due to effectiveness and convenience, cross-subject ITL approaches are commonly researched. Note that PTL usually fulfills the feature extraction and classification altogether when transferring the model between domains. In terms of results, most TL algorithms can achieve good classification performance (over 70%) using no or few labeled training samples from the target domain.

SEMI-SUPERVISED LEARNING

Concepts

Transfer learning can remedy the shortage of the labeled samples from the target domain by importing amounts of labeled data from different source domains. However, it is challenging for TL to cope with huge differences among data available. SSL can make full use of the labeled training samples and unlabeled testing samples from the target subject. Because of fewer data, SSL has less pressure to deal with differences among data than TL.

Semi-supervised learning inherits the merits from supervised learning and unsupervised learning. On the one hand, it is helpful for supervised learning to generate a reliable classifier using amounts of discriminative labeled training samples. Nevertheless, the classification performance of supervised learning decreases sharply with the reduction of labeled training samples. On the other hand, it is of great importance for unsupervised learning to explore the inherent information embedded in amounts of unlabeled testing samples. However, it is difficult to apply unsupervised learning to non-stationary EEG-based BCI. To minimize the calibration procedure, it is crucial for SSL to effectively utilize the labeled training set and unlabeled testing set simultaneously.

In general, SSL methods make additional assumptions on relationship between data distribution and decision function. These include the cluster assumption, the low-density assumption, and the smoothness assumption (Zhu and Goldberg, 2009; Tang et al., 2019). The cluster assumption holds that the samples in the same cluster should belong to the same class. The low-density assumption thinks that the decision boundary should go through the sparse low-density feature space. Both the cluster assumption and the low-density assumption focus on the whole data distribution by adjusting the decision boundary (Fu et al., 2021). However, the smoothness

TABLE 2 | Summary of representative TL approaches.

Pattern	Type	Domain	TL method	Classifier	Result (training/testing)	References
MI	ITL	Subject	RCSP based on dynamic time warping	SVM	CA: over 75% (10+84 × 8/50)	Azab et al. (2020)
MI	ITL	Subject	A manifold embedded knowledge approach	sLDA	BCA: 68.73% (0+144 × 8/144)	Zhang and Wu (2020)
MI	ITL	Subject	Riemannian procrustes analysis	MDM	CA: 80% (50+200 × 49/150)	Rodrigues et al. (2019)
MI	ITL	Subject	Euclidean alignment TL framework	CSP-LDA	BCA: 73.53% (0+144 × 8/144)	He and Wu (2020b)
MI	ITL	Task	Label alignment TL framework	TS-SVM	CA: over 70% (20+1400 × 1/124)	He and Wu (2020a)
SSVEP	FRT	Device	A task-related component analysis approach	Computing the highest correlation coefficients	CA: over 70% (0+320 × 1/320)	Nakanishi et al. (2019)
MI	FRT	Task	Similar MI tasks transfer based on Fisher ratio	SVM	CA: 81.94% (10+120 × 1/110)	Zheng X. et al. (2020)
Drowsiness	PTL	Subject	A cross-subject framework by ranking and fusing different source models		CC: 0.6448 (20+100 × 8/80)	Wei et al. (2018)
Attentive mental state	PTL	Subject	An end-to-end deep CNN approach		CA: 79.26% (20+40 × 119/20)	Fahimi et al. (2019)

assumption pays more attention to the local data distribution. It is considered that the samples close to each other are likely from the same class (Engelen and Hoos, 2020). Based on such assumptions, different SSL models utilize the training and testing sets in different ways.

(1) Based on the low-density assumption, the transductive SVM (TSVM) model seeks the optimal hyperplane which can go through the low-density area and separate both labeled training and unlabeled testing data with maximum margin (Joachims, 1999).
(2) Based on the cluster assumption, the generative model iteratively performs the expectation step and the maximization step to learn the posterior distributions and optimal parameters for the testing samples surrounding the small training set (Tkachenko and Lauw, 2017).
(3) Based on the cluster assumption, the self-training model usually adopts a specific supervised learning approach as the base classifier to iteratively train the training samples and the extended testing samples with high confidence (Chen et al., 2016; Wang et al., 2017). The extended testing samples come from the testing set. The rest of testing set is usually used to estimate the model's classification performance. If the size of the extended testing set is 0, the whole testing set is used to build and evaluate the classifier. Likewise, the co-training model iteratively trains two base supervised learning classifiers using each other's previous classification results (Ren et al., 2014).
(4) Based on the low-density assumption and the smoothness assumption, the graph-based model constructs one or more weighted graphs to analyze the similarity between the training and testing samples and then predicts the classes of the testing samples (Zhao et al., 2015; Gan et al., 2018).

Next, we discuss different SSL models in detail.

TSVM Model

Transductive SVM was firstly proposed for text classification (Joachims, 1999). Recently, TSVM methods have been gradually used in EEG-based BCI, since SVM is much suitable for EEG signals (Liao et al., 2007; Bi et al., 2019; Xu et al., 2019a).

Liao et al. (2007) introduced a TSVM method for reducing the training effort in EEG-based BCI. A non-linear TSVM based Gaussian kernel approach was proposed to find a more complex structure in the feature space. Particularly, the particle swarm optimization method is used to tune the best parameters for non-linear TSVM.

Bi et al. (2019) developed a prior information-based TSVM (PI-TSVM) method in ERP-based driver-vehicle interfaces. The prior information refers to the ratio of positive samples to negative samples. Its value is set to one to two according to the experimental paradigm. Based on the framework of TSVM, a SVM classifier is trained using the initial labeled training set. The unlabeled testing samples are sorted based on their decision scores in a descending order. Then, using the prior information, the first third and the rest of the sorted testing samples are initially marked as positive labels and negative labels, respectively. To seek optimal hyperplane which can go through the low-density area, a pair of testing samples are repeatedly selected to switch their pseudo labels according to a certain criterion.

Compared with SVM, TSVM approaches have better generalization ability in the case of limited training set. However, they are easily affected by outlier testing samples (Li and Zhou, 2015).

Generative Model

To our knowledge, there are few literatures about generative model used in EEG-based BCI.

Based on the cluster assumption, Fu et al. (2021) presented a semi-supervised discriminative rectangle mixture approach in MI-based BCI. It is assumed that the prior distribution of decision boundaries is a Gaussian mixture model. Meanwhile, all samples belonging to a specific cluster are supposed within a corresponding decision rectangle. The expectation maximization algorithm is performed to obtain the optimal model by maximizing a posterior estimate for the posterior distributions.

Self-Training Model and Co-training Model

Both the self-training and co-training models are popular because they can take advantage of one or two supervised classifiers. The self-training model iteratively trains a base supervised classifier using the extended training set, whereas the co-training model iteratively trains two base supervised classifiers marking the labels of the extended testing samples for each other. The extended training set is comprised of the extended testing samples from the testing set and the initial training samples. The framework of self-training model and that of co-training model are shown in **Figures 4A,B**, respectively.

The self-training model can be effectively used for offline and online SSL. In offline SSL, the extended testing samples are obtained *a priori*, whereas in online SSL, they are obtained sequentially and labeled on-the-fly. There are many literatures related to offline SSL, such as TSVM methods, generative methods, and graph-based methods. However, in the online scenario, the self-training methods can expand the training set with newly labeled testing samples and update the classifier along with the adaption of the subject's mental state.

Based on spectral regression kernel discriminant analysis (spectral regression KDA, SRKDA), Nicolas-Alonso et al. (2015) presented an offline SSL version, namely semi-supervised SRKDA (SS-SRKDA), and an online SSL version, namely sequential updating SS-SRKDA (SUSS-SRKDA). As a base classifier, SRKDA can find optimal projection casting KDA into a regression framework and save the computational cost by Cholesky decomposition. Based on the self-training framework, SS-SRKDA iteratively trains the SRKDA classifier with the training set and the testing set until convergence. SUSS-SRKDA sequentially labels the arriving testing samples using the SS-SRKDA classifier.

The total set can also be divided into the training, validation, and testing sets. To better utilize the unlabeled testing samples, a transductive learning with covariate shift-detection (TLCSD) method was proposed in MI-based BCI (Raza et al., 2016). First, an inductive SVM classifier is trained using the initial labeled training set. The validation set is used to obtain the optimal hyperparameters. Each testing sample is labeled by the trained classifier. Then, for each unlabeled testing sample, TLCSD selects its K nearest labeled neighbors and calculates the confidence ratio to judge its usefulness. Finally, both the useful testing sample and its estimated label are added to the existing training set to update the SVM classifier. This process is repeated until all testing samples are processed.

Active learning (AL) is an iterative SSL technique for identifying maximally informative samples. Therefore, it is much suitable for selecting testing samples with high confidence. Marathe et al. (2016) presented an AL implementation in a rapid serial visual presentation (RSVP) paradigm. The proposed approach uses a Query-by-Committee (QBC) framework with a heterogeneous ensemble of popular models, including hierarchical discriminant component analysis (HDCA), CSP+BLDA (CSP-BLDA), xDawn+Bayesian LDA (XD-BLDA). First, the HDCA, CSP-BLDA, and XD-BLDA models are separately trained using the training set. Then, the area-under-curve (AUC) is calculated to measure model efficacy using the validation set. The testing samples are labeled by different models. For each testing sample, its aggregate confidence is calculated based on the distance of each model's score from the discriminating boundary. The testing samples with the lowest aggregate confidence are chosen to be informative trials and sent to the oracle for labeling. The oracle is assumed to be the user himself.

For co-training model, since the two base classifiers give different separate planes, they are good complements to each other.

Ren et al. (2014) proposed a co-training method based on a biomimetic pattern recognition (BPR) classifier and a sparse presentation (SP) classifier for MI-based BCI. As the base classifiers, both BPR and SR are computationally simple due to few parameters to be tuned. Moreover, they provide reasonably different separate planes, which is important for co-training model. BPR is based on "coverage." It consists of multiple neurons trained by training data. While SR is based on the compressed sensing theory, which takes advantage of the redundancy in many signals. For different base classifiers, the distance to one class and the residuals of one class are simultaneously considered to select high confident extended testing samples.

Meng et al. (2014) presented a co-training approach combining a LDA classifier with a BLDA classifier. The CSP feature extraction and the co-training classifier are performed jointly and iteratively (Meng et al., 2014). LDA can give good classification performance and fast computational cost simultaneously. While BLDA can give better generalization ability as the Bayesian version of LDA. For different classifiers, the extended testing samples with largest absolute values of classification scores are selected for each other.

Although both the self-training and co-training models are simple and effective, they are susceptible to the incorrect labels of the extended testing samples with high confidence during iterations.

Graph-Based Model

The graph-based model can construct one or more weighted graphs to reveal the manifold structure behind the training set or the total set. The graph Laplacian is always embedded in the objective function of the SSL classifier.

A semi-supervised quantization approach based on the cartesian K-means (SSCK) method was proposed for MI and emotion EEG classification, where the graph Laplacian was built based on the labeled training set and added in the objective function as a manifold regularization term (Liu M. et al., 2020). $L = D - W$ is the Laplacian matrix. Based on the cluster assumption, the entry W_{ij} of the similarity matrix W is assigned a large weight when the i-th and j-th labeled training samples are in the same cluster, otherwise it is assigned a small weight. D is a diagonal matrix with its diagonal elements $D_{ii} = \sum_j W_{ij}$. To get more discriminative

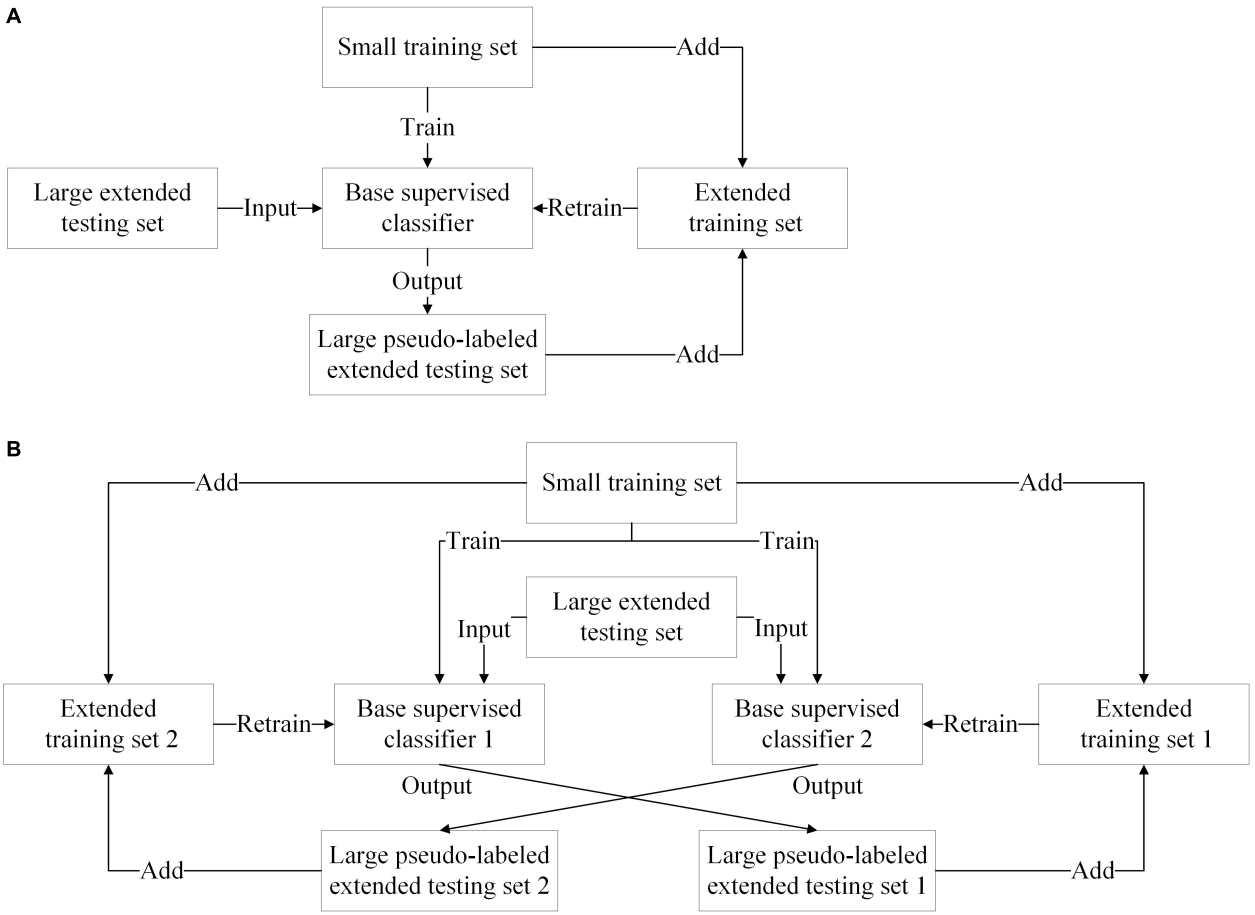


FIGURE 4 | Self-training and co-training models. **(A)** The framework of self-training model. (**B**) The framework of co-training model.

SSCK model, the following objective function is iteratively optimized:

$$\underset{R,\ C,\ B,Y^{(l)},\mu}{\operatorname{argmin}} (\parallel R^T X_{\mu}^{(l)} - CY^{(l)} \parallel_F^2 + \eta \parallel R^T X_{\mu}^{(u)} - CB \parallel_F^2 + \lambda tr(Y^{(l)} L (Y^{(l)})^T)), \tag{11}$$

where the first and second terms are respectively employed to reduce the squared distortion errors in the transformed training features $X_{\mu}^{(l)} \equiv X^{(l)} - \mu\ (\mathbf{1}^{(l)})^{\mathbf{T}}$ and the transformed testing features $X_{\mu}^{(u)} \equiv X^{(u)} - \mu\ (\mathbf{1}^{(u)})^{\mathbf{T}}$ by the K-means clustering with respect to the rotation matrix R and the cluster center C. $X^{(l)}$ and $X^{(u)}$ are the original training and testing features, respectively. μ is the mean value vector of input data after each iteration. The third term is the supervised Laplacian regularization term. $Y^{(l)}$ and B are separately the training label matrix and the testing label matrix. $tr(\cdot)$ performs the trace computation. η and λ are the trade-off parameters.

Based on the extreme learning machine (ELM) and deep architecture, a hierarchical semi-supervised ELM (SS-ELM) method was presented for MI task recognition, where the graph Laplacian was calculated using all samples (She et al., 2019a). Firstly, high-level features $G_{\tilde{k}}$, $T_{\tilde{k}}$, and $H_{\tilde{k}}$ are separately extracted from l labeled training samples, u unlabeled testing samples, and $(l+u)$ total samples by deep architecture, where $\tilde{k}$ is the number of hidden layers. Then, these features are input to the SS-ELM classifier to optimize the objective function as follows:

$$\underset{\beta}{\operatorname{argmin}}(\parallel \beta \parallel_2^2 + \tilde{C} \parallel G_{\tilde{k}}\beta - \tilde{Y} \parallel_2^2 + \lambda \beta^T H_{\tilde{k}}{}^T \tilde{L} H_{\tilde{k}}\beta), \tag{12}$$

where the output weights β can be calculated by Moore-Penrose principle, $\tilde{C}$ is a penalty coefficient on the training errors, $\tilde{Y}$ denotes the labels of training data, λ is a trade-off parameter. $\tilde{L} = \tilde{D} - \tilde{W}$ is the graph Laplacian, $\tilde{D}$ denotes a diagonal matrix, $\tilde{W}_{ij}$ is the element of the similarity matrix $\tilde{W}$ $(1 \leq i,\ j \leq l+u)$. Finally, $\tilde{Y}_{predict} = \beta T_{\tilde{k}}$ is the labels of testing samples. The proposed approach takes advantage of ELM, deep architecture, and graph-based model.

A k-nearest neighbor graph is usually included in the Laplacian matrix to define the relationship among the nearby data points (She et al., 2019b; Peng et al., 2020).

Based on the graph label propagation algorithm and the broad learning system (BLS), a graph-based semi-supervised BLS was

proposed in MI-based BCI system (She et al., 2019b). On the one hand, the label information can be smoothed over the graph by the graph label propagation algorithm. On the other hand, BLS has a simpler structure, fewer layers, and a shorter training time than DL. First, the proposed algorithm constructs a k-nearest weight matrix $\hat{W}$ as follows:

$$\hat{W}_{ij} = \begin{cases} e^{-\|x_i - x_j\|^2/\sigma^2} & \text{if } \begin{cases} x_i \in N(x_j) \\ x_j \in N(x_i) \end{cases} \\ 0 & \text{otherwise} \end{cases}, \quad (13)$$

where $N(\cdot)$ is the set of k nearest neighbors, σ is a scaling parameter of Gaussian function, x_i and x_j belong to the total set. To obtain the optimal soft label matrix F, the objective function is optimized as below:

$$\underset{F}{\operatorname{argmin}} \left(\sum_{i,j} \| F_i - F_j \|^2 \hat{W}_{ij} + \hat{\mu} \sum_{i,j} \| x_i - x_j \|^2 \hat{W}_{ij} \right), \quad (14)$$

where the first and second terms constrain that the similar trials have similar labels and features, respectively. $\hat{\mu}$ is used to balance the roles of label space F and feature space x. After optimization, the testing samples are labeled. Then, all features with their labels are successively fed into the feature layer, the enhancement layer, and the output layer of BLS to obtain their prediction labels.

As mentioned above, more than one graphs can be exploited to represent the relationship among all samples.

Combining a label-consistency graph (LCG) and a sample-similarity graph (SSG), a balanced graph-based regularized SS-ELM for MI EEG classification was presented (She et al., 2021). The SSG is the traditional graph Laplacian which reveals the similarity between the training and extended testing data without using the label information (She et al., 2019a). Moreover, to develop the class similarity among the training and extended testing data, the LCG defines a new adjacent graph matrix as follows:

$$W'_{ij} = \begin{cases} 1/N_{t-th} & \text{if both } h_i \text{ and } h_j \text{ belong to the } t-\text{th class} \\ 0 & \text{otherwise} \end{cases}, \quad (15)$$

where h_i and h_j are separately the mapping vectors of x_i and x_j represented by the hidden layer of ELM, N_{t-th} denotes the number of the t-th class samples. The pseudo labels of all extended testing samples are obtained by traditional ELM classifier. Then, like the traditional graph Laplacian, the LCG is the new Laplacian matrix $L' = D' - W'$, where D' is a diagonal matrix with its diagonal elements $D'_{ii} = \sum_j W'_{ij}$. The balanced graph is the linear combination of LCG and SSG with different weights. Finally, such balanced graph is embedded in the objective function of the SS-ELM classifier.

The summary of some representative SSL approaches is shown in **Table 3**.

As shown in **Table 3**, most SSL algorithms are applied in the MI-based BCI. Both the self-training and graph-based models are popular. The reason is that the former can take advantage of different base supervised classifiers and then can be easily realized. The latter can directly reveal the sample-similarity or label-similarity among data available. In terms of result, most SSL classifiers can obtain good classification accuracies for two-class problem with the help of a few training samples, the extended testing set or validation set. For multi-class BCI tasks, SSL classifiers can get good kappa values, too.

A COMBINATION OF TRANSFER LEARNING AND SEMI-SUPERVISED LEARNING

Concepts

As mentioned above, TL often performs in the feature extraction stage and aims to extract more distinctive target features by

TABLE 3 | Summary of representative SSL approaches.

Pattern	Type	Feature extraction	Classifier	Result [training/validation (extended testing)/testing]	References
ERP	TSVM	Variational autoencoder	A prior information based TSVM method	CA: 85% (30/0/240)	Bi et al. (2019)
MI	Self-training	CSP	Online spectral regression kernel discriminant analysis	CA: 77% (144/0/144)	Nicolas-Alonso et al. (2015)
MI	Self-training	CSP	Transductive learning with covariate shift-detection	CA: 69.72% (240/160/320)	Raza et al. (2016)
RSVP	Self-training	Active learning with an ensemble of models, including HDCA, CSP-BLDA, and XD-BLDA		AUC: over 0.9 (50%/10%/40%)	Marathe et al. (2016)
MI	Co-training	CSP	Both BRP and sparse presentation are assigned as base classifiers	CA: over 70% (14/126/140)	Ren et al. (2014)
MI	Graph	CNN	A semi-supervised quantization approach based on the cartesian K-means method	CA: 76.86% (28/0/260)	Liu M. et al. (2020)
MI	Graph	CSP	A graph-based semi-supervised broad learning system	Kappa value: 0.525 (288/0/288)	She et al. (2019b)
MI	Graph	CSP	A balanced graph-based regularized semi-supervised ELM	Kappa value: 0.660 (144/144/288)	She et al. (2021)

transferring the labeled samples from the source domains, whereas SSL often performs in the classification stage and aims to utilize the labeled and unlabeled samples from the target domain. Recently, a combination of TL and SSL has gained an increasing attention (Jiang et al., 2017; Wu, 2017a,b; Zhao et al., 2019; Zhou et al., 2021). To our best knowledge, it occurs in two ways:

(1) A combination of cross-session TL and SSL. For cross-session TL, the samples from the source and target sessions are usually labeled and unlabeled, respectively. They come from the same subject. Therefore, the samples used in cross-session TL are the same to those used in SSL. However, cross-session TL focuses on reducing the discrepancy in data distribution between different sessions, SSL focuses on utilizing the useful information contained in the unlabeled testing data.
(2) A combination of cross-subject TL and SSL. For cross-subject TL, it is a great challenge to deal with high inter-subject variability. For SSL, it is hard to collect amounts of labeled training samples and unlabeled testing samples due to high inter-session variability. After combination, cross-subject TL can relieve the pressure of limited samples for SSL, and SSL can reduce the variability for cross-subject TL since the inter-session variability is generally smaller than inter-subject variability.

Next, we present different kinds of approaches related to the combination of TL and SSL.

A Combination of Cross-Session TL and SSL

To achieve epileptic seizure classification, Jiang et al. (2017) integrated TL, SSL, and a Takagi-Sugeno-Kang (TSK) fuzzy system. As the base classifier, the TSK fuzzy system can increase the model interpretability for medical diagnostics. Cross-session TL can minimize the differences between the source and target sessions. Based on the cluster assumption, SSL is designed for label clustering. The integrated model aims to seek optimal projection P_g used in the decision function. The objective function of the proposed model consists of three parts:

Part 1:

To utilize the labeled training samples from the source session, the objective function used for the TSK fuzzy model is defined as follows:

$$\underset{P_g}{\operatorname{argmin}}\left(\frac{1}{2}tr\left((P_g{}^T x_{g,s}-y_s)(P_g{}^T x_{g,s}-y_s)^T\right)+\frac{\lambda_1}{2}tr(P_g{}^T P_g)\right), \tag{16}$$

where $(x_{g,s}, y_s)$ denotes the new labeled set constructed by Fuzzy C-means (FCM) clustering, λ_1 denotes a trade-off parameter balancing the complexity of the model and the tolerance of error. The two terms are used to regularize the TSK fuzzy model.

Part 2:

To reduce the differences between the source session (labeled set) and the target session (unlabeled set), the objective function used for cross-session TL is optimized as below:

$$\underset{P_g}{\operatorname{argmin}}(\lambda_2 tr(P_g{}^T \Omega P_g)), \tag{17}$$

where

$$\Omega=\frac{1}{{N_T}^2}x_{g,T}[\mathbf{1}]^{N_T\times N_T}x_{g,T}^T+\frac{1}{{N_S}^2}x_{g,s}[\mathbf{1}]^{N_S\times N_S}x_{g,s}^T - \frac{1}{N_S N_T}(x_{g,T}[\mathbf{1}]^{N_T\times N_S}x_{g,s}^T+x_{g,s}[\mathbf{1}]^{N_S\times N_T}x_{g,T}^T). \tag{18}$$

λ_2 is a regularization parameter. N_S is the size of source session. N_T is the size of target session. $x_{g,T}$ denotes the new unlabeled testing set from target session constructed by FCM. The term in Equation (17) is used to minimize the projected squared maximum mean discrepancy (MMD) distance between different sessions.

Part 3:

To label the unlabeled testing samples from the target session, the objective function used for SSL is computed as follows:

$$\underset{P_g}{\operatorname{argmin}}\left(\lambda_3\sum_{j=1}^{\hat{C}}\sum_{i=1}^{N_T}\widehat{\mu}_{i,j}\parallel P_g{}^T x_{gi,T}-\theta_j\parallel^2\right), \tag{19}$$

where λ_3 is a regularization term, $\hat{C}$ is the number of clusters, $\widehat{\mu}_{i,j}$ denotes the label membership of the i-th unlabeled sample $x_{gi,T}$ belonging to the j-th cluster, and $\theta_j=[0,\cdots,0,1_{j\text{th}},0,\cdots,0]^T$ denotes a label vector of the j-th cluster. The term is used to assign the data close to each other to the same class. Because the optimal P_g is calculated using the labeled samples from the source session and the unlabeled samples from the target session according to the three parts mentioned above, the objective function is optimized in a semi-supervised way.

The objective function in Equations (16–19) is iteratively optimized to obtain the optimal projection P_g. The proposed model increases the classification performance by integrating the supervised TSK fuzzy system, cross-session TL, and SSL.

TABLE 4 | Summary of representative TL+SSL approaches.

Pattern	Type	Feature extraction	Classifier	Result (training/testing)	References
Epileptic	Cross-session TL+SSL	Wavelet packet decomposition/Short time Fourier transform/Kernel PCA	An integrated model combing TL, SSL and the TSK fuzzy system	CA: over 95% (0+150 × 1/50)	Jiang et al. (2017)
MI	Cross-subject TL+SSL	EEGLAB	A weighted adaptation regularization with source domain selection	BCA: about 70% (80+270 × 13/190)	Wu (2017b)
MI	Cross-subject TL+SSL	CSP filters selection based on a clustering method	Semi-supervised SVM	CA: about 70% (40+280 × 4/240)	Zhao et al. (2019)

A Combination of Cross-Subject TL and SSL

Compared with the traditional TL methods, the combination of cross-subject TL and SSL adds the unlabeled testing samples from the target subject to build a more subject-specific model.

Originating from a weighted adaptation regularization (wAR) method (Wu et al., 2016), an offline wAR with source domain selection (wARSDS) approach was proposed (Wu, 2017b).

First, the pseudo labels of testing samples from the target subject are initially estimated by the wAR classifier trained using the training samples from the source and target subjects.

Then, the labeled training samples and pseudo-labeled testing samples from the target subject are used to calculate the class centroids of the target subject. To save the computational time, based on *k*-means clustering, most suitable source subjects are selected, whose class distances to the class centroids of the target subject are comparably smaller.

Finally, using all samples from the target subject and selected source subjects, the loss function minimization, the marginal probability distribution adaptation, and the conditional probability distribution adaptation are iteratively performed to update the wARSDS classifier.

Zhao et al. (2019) transferred common spatial filters with SSL for MI BCI. First, for each subject, 30 groups of filters are calculated by CSP after randomly selecting 60 samples per class for 30 times. The representative CSP filters are selected from all filters available based on the clustering method.

Then, all new CSP features from the source and target subjects are generated by projecting these representative CSP filters into the original pre-processed samples.

Finally, the semi-supervised SVM classifier makes full use of the labeled training features from the source subjects and all features from the target subject by optimizing the expectation risk, Tikhonov regularization term, and the manifold regularization. A graph Laplacian with Gaussian kernel function is embedded in the objective function using all features available.

In **Table 4**, some representative TL+SSL models are briefly listed.

In the fifth column of **Table 4**, like **Table 2**, 'Training' refers to the size of all training sets from the target and source domains. Like **Table 3**, 'Testing' refers to the size of testing set from the target domain. As shown in **Table 4**, a combination of TL and SSL is effective since their classification performance can also reach 70% on average using few training samples from the target domain. Therefore, the combination of TL and SSL will be promising in future.

DISCUSSION

Based on many papers surveyed herein, we briefly summarize various signal processing approaches to reduce calibration effort in EEG-based BCI.

(1) In terms of EEG-based BCI
Compared with the traditional EEGs, such as SSVEP and P300, MI EEG is more widely studied in the small training set scenario. The reason is that there are numerous MI datasets available. Moreover, MI is harder to control for both healthy and unhealthy subjects. It is more urgent for MI to minimize calibration time. Thus, many base feature learning and classifier learning methods applied to MI are always referred, such as CSP, LDA, SVM, MDM and so on. Additionally, drive drowsiness and emotion state estimations also need to shorten calibration effort for their complex BCI tasks. Similar methods can be used to solve this problem, such as ensemble models, CNN, etc.

(2) In terms of TL
Compared with SSL, TL is more popular regardless of high inter-domain variability. The reason is that TL provides more sufficient samples from the source domains than SSL and promotes the development of generic model.
As shown in **Table 2**, Most TL approaches can reach satisfied classification performance (over 70%) even if the calibration time is reduced or suppressed. For a calibration-free TL model, unlabeled samples from the target domain can also build a subject-specific model.
Currently, most TL approaches aim to adapt the source domains to the small target domain. However, if the distribution of the small target domain is not discriminable, the classification performance might decrease sharply after such adaptation. To our best knowledge, few literatures cope with this problem. In our opinion, flexible measures should be taken according to the discriminability of the small target domain. Moreover, most TL approaches are not suitable for online BCI since they are designed offline.
Data alignment-based instance transfer is very promising since it can make aligned covariance matrices from different domains comparable. However, high dimensionality of covariance matrix may lead to curse of dimensionality. Therefore, channel selection is a basic problem for data alignment-based TL. Unsupervised PCA and supervised CSP can be used to cope with this problem.
Deep learning is highly effective using the big training set. However, it cannot be directly used in the small training set scenario. Thus, in the PTL, DL can build a pre-trained model using amounts of samples from source subjects and then adapt the model for the new target subject. CNN is widely used in EEG-based BCI.

(3) In terms of SSL
Although few SSL literatures are proposed to shorten the calibration procedure. SSL classifiers still outstand themselves by discovering the useful information embedded in the labeled and unlabeled samples. In EEG-based BCI, the self-training model can be easily achieved by using the well-known supervised model as the base classifier. Additionally, the graph-based model is popular by providing the overall view based on the total samples.
As shown in **Table 3**, compared with TL approaches, most SSL approaches use more training samples from the target domain to obtain good classification performance. Therefore, it is further testified that the number of samples

available is extremely important to reduce calibration time and achieve good performance.

Most SSL classifiers surveyed in this paper are designed offline. However, the EEG-based BCI is fundamentally online. Self-training SSL model is suitable for the online scenario. It should be combined with other SSL models to adjust their classifiers in real time.

(4) In terms of the combination of TL and SSL

The combination of TL and SSL is emerging and promising since it not only relieves the pressure of limited sample, but also brings more subject-specific samples. It should be further explored to increase its effectiveness.

As shown in **Table 3**, although few approaches relate to such combination, they still achieve good performance (over 70%) with few samples from target domain by taking advantage of TL and SSL.

(5) In terms of all signal processing approaches

To reduce the calibration time, we focus on the signal processing approaches in this paper. Nevertheless, the BCI performance also relies on the subject, the BCI paradigm, and so on. So far, fewer literatures discuss them altogether. In our opinion, different subjects have extremely different performance in EEG-based BCI. Therefore, different signal processing approaches should be designed for different subjects. In addition, more friendly BCI paradigm should be used for better experimental experience.

CONCLUSION

In this study, we have surveyed the signal processing approaches to minimize the calibration procedure in EEG-based BCI. The numerous approaches in this paper can be divided into three main categories: TL, SSL, and a combination of TL and SSL.

We can draw the following conclusions, which may inspire future research directions:

(1) It is more difficult to decode MI EEG than SSVEP, P300, etc. MI EEG-based BCI should be further studied to speed up its application.

(2) TL is instrumental to reduce the calibration time. Data alignment-based TL is a good choice to relieve the differences between domains. TL can be combined with SSL or DL to achieve a calibration-free model.

(3) SSL can build a more subject-specific model than TL. Currently, offline SSL classifiers have been paid more attention than online SSL classifiers. In terms of online BCI control, online SSL classifiers should be further researched.

(4) The performance of EEG-based BCI is closely related to the signal processing approaches, the BCI subject, the paradigm, and so on. Most signal processing approaches in this paper are evaluated offline based on the publicly available BCI datasets. Future signal processing approaches should further adapt to the mental state change of the BCI subject in real time and work using a more friendly paradigm.

AUTHOR CONTRIBUTIONS

YX designed the study. XH and YX wrote the manuscript. JH and WY collected the relevant literatures. HY and RH prepared the figures. SW reviewed and edited the manuscript. All authors read and approved the submitted manuscript.

FUNDING

This work was partially funded by Science and Technology Research Project of Jiangxi Provincial Department of Education (Grant No. GJJ200318), National Natural Science Foundation of China (Grant Nos. 61861021 and 61762048), and Natural Science Foundation of Jiangxi Province (Grant No. 20192BAB206032).

ACKNOWLEDGMENTS

The authors would like to thank the referees for their constructive recommendations, which help to improve the manuscript substantially.

REFERENCES

Adair, J., Brownlee, A., Daolio, F., and Ochoa, G. (2017). "Evolving training sets for improved transfer learning in brain computer interfaces," in *Proceedings of the 2017 International Workshop on Machine Learning, Optimization, and Big Data (Cham)*, (Cham), 186–197. doi: 10.1007/978-3-319-72926-8_16

Alamgir, M., Grosse-Wentrup, M., and Altun, Y. (2010). "Multitask learning for brain-computer interfaces," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (Italy)*, (Italy), 17–24.

Azab, A., Ahmadi, H., Mihaylova, L., and Arvaneh, M. (2020). Dynamic time warping-based transfer learning for improving common spatial patterns in brain-computer interface. *J. Neural Eng.* 17:016061. doi: 10.1088/1741-2552/ab64a0

Azab, A., Mihaylova, L., Ang, K., and Arvaneh, M. (2019). Weighted transfer learning for improving motor imagery-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 7, 1352–1357. doi: 10.1109/TNSRE.2019.2923315

Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a riemannian-based kernel for BCI applications. *Neurocomputing* 112, 172–178. doi: 10.1016/j.neucom.2012.12.039

Bi, L., Zhang, J., and Lian, J. (2019). EEG-based adaptive driver-vehicle interface using variational autoencoder and PI-TSVM. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2025–2033. doi: 10.1109/TNSRE.2019.2940046

Bird, J. J., Kobylarz, J., Faria, D. R., Ekart, A., and Ribeiro, E. P. (2020). Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG. *IEEE Access* 8, 54789–54801. doi: 10.1109/ACCESS.2020.2979074

Blankertz, B., Dornhege, G., and Krauledat, M. (2007). The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *Neuroimage* 37, 539–550. doi: 10.1016/j.neuroimage.2007.01.051

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP componentsa tutorial. *NeuroImage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Blankertz, B., Müller, K.-R., Krusien-ski, D., Schalk, G., Wolpaw, J. R., Schlögl, A., et al. (2006). The BCI competition III: validating alternative approachs to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159. doi: 10.1109/TNSRE.2006.875642

Cao, L., Chen, S., Jia, J., Fan, C., Wang, H., and Xu, Z. (2021). An inter- and intra-subject transfer calibration scheme for improving feedback performance of sensorimotor rhythm-based BCI rehabilitation. *Front. Neurosci.* 14:629572. doi: 10.3389/fnins.2020.629572

Chai, R., Ling, S. H., Hunter, G. P., Tran, Y., and Nguyen, H. T. (2014). Brain–computer interface classifier for wheelchair commands using neural network with fuzzy particle swarm optimization. *IEEE J. Biomed. Health Informat.* 18, 1614–1624. doi: 10.1109/JBHI.2013.2295006

Chai, R., Ling, S. H., San, P. P., Naik, G. R., Nguyen, T. N., Tran, Y., et al. (2017). Improving EEG-based driver fatigue classification using sparse-deep belief networks. *Front. Neurosci.* 11:103. doi: 10.3389/fnins.2017.00103

Chen, M., Tan, X., and Zhang, L. (2016). An iterative self-training support vector machine algorithm in brain-computer interfaces. *Intell. Data Anal.* 49, 67–82. doi: 10.3233/IDA-150794

Cheng, M., Lu, Z., and Wang, H. (2017). Regularized common spatial patterns with subject-to-subject transfer of EEG signals. *Cogn. Neurodyn.* 11, 173–181. doi: 10.1007/s11571-016-9417-x

Chiang, K. J., Wei, C. S., Nakanishi, M., and Jung, T. P. (2019). "Cross-subject transfer learning improves the practicality of real-world applications of brain-computer interfaces," in *Proceedings of the 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, (San Francisco CA), 424–427. doi: 10.1109/NER.2019.8716958

Dai, M., Wang, S., Zheng, D., Na, R., and Zhang, S. (2019). Domain transfer multiple kernel boosting for classification of EEG motor imagery signals. *IEEE Access* 7, 49951–49960. doi: 10.1109/ACCESS.2019.2908851

Dai, M., Zheng, D., Liu, S., and Zhang, P. (2018). Transfer kernel common spatial patterns for motor imagery brain-computer interface classification. *Comput. Math. Method. Med.* 2018:9871603. doi: 10.1155/2018/9871603

Dose, H., Møller, J. S., Iversen, H. K., and Puthusserypady, S. (2018). An end-to-end deep learning approach to MI-EEG signal classification for BCIs. *Expert Syst. Appl.* 114, 532–542. doi: 10.1016/j.eswa.2018.08.031

Engelen, J. V., and Hoos, H. H. (2020). A survey on semi-supervised learning. *Mach. Learn.* 109, 373–440. doi: 10.1007/s10994-019-05855-6

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T., Ang, K. K., and Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *J. Neural Eng.* 16:026007. doi: 10.1088/1741-2552/aaf3f6

Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Fu, R., Li, W., Chen, J., and Han, M. (2021). Recognizing single-trial motor imagery EEG based on interpretable clustering method. *Biomed. Signal Proces.* 63:102171. doi: 10.1016/j.bspc.2020.102171

Gan, H., Li, Z., Wu, W., Luo, Z., and Huang, R. (2018). Safety-aware graph-based semi-supervised learning. *Expert Sys. Appl.* 107, 243–254. doi: 10.1016/j.eswa.2018.04.031

He, H., and Wu, D. (2020a). Different set domain adaptation for brain-computer interfaces: a label alignment approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1091–1108. doi: 10.1109/TNSRE.2020.2980299

He, H., and Wu, D. (2020b). Transfer learning for brain-computer interfaces: a Euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* 67, 399–410. doi: 10.1109/TBME.2019.2913914

Hehenberger, L., Kobler, R. J., Lopes-Dias, C., Srisrisawang, N., Tumfart, P., Uroko, J. B., et al. (2021). Long-term mutual training for the CYBATHLON BCI race with a tetraplegic pilot: a case study on inter-session transfer and intra-session adaptation. *Front. Hum. Neurosci.* 15:635777. doi: 10.3389/fnhum.2021.635777

Hohyun, C., Minkyu, A., Sangtae, A., Moonyoung, K., and Sung, C. J. (2017). EEG datasets for motor imagery brain computer interface. *Gigascience* 6:gix034. doi: 10.1093/gigascience/gix034

Horki, P., Klobassa, D. S., Pokorny, C., and Müller-Putz, G. R. (2015). Evaluation of healthy EEG responses for spelling through listener-assisted scanning. *IEEE J. Biomed. Health Informat.* 19, 29–36. doi: 10.1109/JBHI.2014.2328494

Hossain, I., Khosravi, A., Hettiarachchi, I., and Nahavandi, S. (2018). "Calibration time reduction using subjective features selection based transfer learning for multiclass BCI," in *Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (Miyazaki), 491–498. doi: 10.1109/SMC.2018.00093

Huang, C., Zong, Y., Chen, J., Liu, W., LIore, J., and Mukherjee, M. (2021). A deep segmentation network of stent structs based on IoT for interventional cardiovascular diagnosis. *IEEE Wirel. Commun.* 28, 36–43. doi: 10.1109/MWC.001.2000407

Jiang, Y., Wu, D., Deng, Z., Qian, P., Wang, J., Wang, G., et al. (2017). Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 2270–2284. doi: 10.1109/TNSRE.2017.2748388

Jiao, Y., Zhang, Y., Chen, X., Yin, E., Jin, J., and Wang, X. (2019). Sparse group representation model for motor imagery EEG classification. *IEEE J. Biomed. Health* 23, 631–641. doi: 10.1109/JBHI.2018.2832538

Jin, J., Li, S., Miao, Y., Liu, C., and Wang, X. (2020). The study of generic model set for reducing calibration time in P300-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 3–12. doi: 10.1109/TNSRE.2019.2956488

Jin, Y., Mousavi, M., and de Sa, V. R. (2018). "Adaptive CSP with subspace alignment for subject-to-subject transfer in motor imagery brain-computer interfaces," in *Proceedings of the 2018 6th International Conference on Brain-Computer Interface (BCI)*, (GangWon), 1–4. doi: 10.1109/IWW-BCI.2018.8311494

Joachims, T. (1999). "Transductive inference for text classification using support vector machines," in *Proceedings of the 16th International Conference on Machine Learning (ICML'99)*, (Bled), 200–209.

Kang, H., Nam, Y., and Choi, S. (2009). Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Process. Lett.* 16, 683–686. doi: 10.1109/LSP.2009.2022557

Keng, A. K., Yang, C. Z., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* 6:39. doi: 10.3389/fnins.2012.00039

Khalaf, A., and Akcakaya, M. (2020). A probabilistic approach for calibration time reduction in hybrid EEG–fTCD brain–computer interfaces. *Biomed. Eng. Online* 19, 295–314. doi: 10.1186/s12938-020-00765-4

Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J., and Birbaumer, N. (2001). Brain-computer communication: unlocking the locked in. *Psychol. Bull.* 127, 358–375. doi: 10.1037/0033-2909.127.3.358

Kumar, S. U., and Inbarani, H. H. (2017). PSO-based feature selection and neighborhood rough set-based classification for BCI multiclass motor imagery task. *Neural Comput. Appl.* 28, 3239–3258. doi: 10.1007/s00521-016-2236-5

Li, Y., and Zhou, Z. (2015). Towards making unlabeled data never hurt. *IEEE Trans. Pattern Anal.* 37, 175–188. doi: 10.1109/TPAMI.2014.2299812

Liao, X., Yao, D., and Li, C. (2007). Transductive SVM for reducing the training effort in BCI. *J. Neural Eng.* 4, 246–254. doi: 10.1088/1741-2560/4/3/010

Liu, M., Zhou, M., Zhang, T., and Xiong, N. (2020). Semi-supervised learning quantization algorithm with deep features for motor imagery EEG recognition in smart healthcare application. *Appl. Soft Comput.* 89:106071. doi: 10.1016/j.asoc.2020.106071

Liu, Y., Lan, Z., Cui, J., Sourina, O., and Müller-Wittig, W. (2020). Inter-subject transfer learning for EEG-based mental fatigue recognition. *Adv. Eng. Inform.* 46:101157. doi: 10.1016/j.aei.2020.101157

Lotte, F. (2015). Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proc. IEEE* 103, 871–890. doi: 10.1109/JPROC.2015.2404941

Ma, J., Zhang, Y., Cichocki, A., and Matsuno, F. (2015). A novel EOG/EEG hybrid human–machine interface adopting eye movements and erps: application to robot control. *IEEE Trans. Biomed. Eng.* 62, 876–889. doi: 10.1109/TBME.2014.2369483

Marathe, A. R., Lawhern, V. J., Wu, D., Slayback, D., and Lance, B. J. (2016). Improved neural signal classification in a rapid serial visual presentation task using active learning. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 333–343. doi: 10.1109/TNSRE.2015.2502323

Meng, J., Sheng, X., Zhang, D., and Zhu, X. (2014). Improved semisupervised adaptation for a small training dataset in the brain–computer interface. *IEEE J. Biomed. Health Infor.* 18, 1461–1472. doi: 10.1109/JBHI.2013.2285232

Ming, Y., Pelusi, D., Fang, C., Prasad, M., Wang, Y., Wu, D., et al. (2018). EEG data analysis with stacked differentiable neural computers. *Neural Comput. Appl.* 32, 7611–7621. doi: 10.1007/s00521-018-3879-1

Moakher, M. (2006). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* 26, 735–747. doi: 10.1137/S0895479803436937

Muhl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain Comput. Interfaces* 1, 66–84. doi: 10.1080/2326263X.2014.912881

Nakanishi, M., Wang, Y. T., Wei, C. S., Chiang, K. J., and Jung, T. P. (2019). Facilitating calibration in high-speed BCI spellers via leveraging cross-device shared latent responses. *IEEE Trans. Biomed. Eng.* 4, 1105–1113. doi: 10.1109/TBME.2019.2929745

Nicolas-Alonso, L. F., Corralejo, R., Gomez-Pilar, J., Álvarez, D., and Hornero, R. (2015). Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain-computer interfaces. *Neurocomputing* 159, 186–196. doi: 10.1016/j.neucom.2015.02.005

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Peng, Y., Li, Q., Kong, W., Qin, F., and Zhang, J. (2020). A joint optimization framework to semi-supervised RVFL and ELM networks for efficient data classification. *Appl. Soft Comput.* 97:106756. doi: 10.1016/j.asoc.2020.106756

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain computer communication. *Proc. IEEE* 89, 1123–1134. doi: 10.1109/5.939829

Qi, H., Xue, Y., Xu, L., Cao, Y., and Jiao, X. (2018). A speedy calibration method using Riemannian geometry measurement and other-subject samples on a P300 speller. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 602–608. doi: 10.1109/TNSRE.2018.2801887

Rao, R. (2010). Brain-computer interfacing. *IEEE Signal Proc. Mag.* 27, 177–209. doi: 10.1017/CBO9781139032803

Raza, H., Cecotti, H., Li, Y., and Prasad, G. (2016). Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface. *Soft Comput.* 20, 3085–3096. doi: 10.1007/s00500-015-1937-5

Ren, Y., Wu, Y., and Ge, Y. (2014). A co-training algorithm for EEG classification with biomimetic pattern recognition and sparse representation. *Neurocomputing* 137, 212–222. doi: 10.1016/j.neucom.2013.05.045

Riccio, A., Esimione, L., Eschettini, F., Epizzimenti, A., Inghilleri, M., Belardinelli, M. E., et al. (2013). Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Front. Hum. Neurosci.* 7:732. doi: 10.3389/fnhum.2013.00732

Rodrigues, P. L. C., Jutten, C., and Congedo, M. (2019). Riemannian procrustes analysis: transfer learning for brain–computer interfaces. *IEEE Trans. Biomed. Eng.* 66, 2390–2401. doi: 10.1109/TBME.2018.2889705

Sannelli, C., Vidaurre, C., Müller, K., and Blankertz, B. (2016). Ensembles of adaptive spatial filters increase BCI performance: an online evaluation. *J. Neural. Eng.* 13:46003. doi: 10.1088/1741-2560/13/4/046003

She, Q., Hu, B., Luo, Z., Nguyen, T., and Zhang, Y. (2019a). A hierarchical semi-supervised extreme learning machine method for EEG recognition. *Med. Biol. Eng. Comput.* 57, 147–157. doi: 10.1007/s11517-018-1875-3

She, Q., Zhou, Y., Gan, H., Ma, Y., and Luo, Z. (2019b). Decoding EEG in motor imagery tasks with graph semi-supervised broad learning. *Electronics* 8:1273. doi: 10.3390/electronics8111273

She, Q., Zou, J., Luo, Z., Nguyen, T., Li, R., and Zhang, Y. (2020). Multi-class motor imagery EEG classification using collaborative representation-based semi-supervised extreme learning machine. *Med. Biol. Eng. Comput.* 2020, 2119–2130. doi: 10.1007/s11517-020-02227-4

She, Q., Zou, J., Meng, M., Fan, Y., and Luo, Z. (2021). Balanced graph-based regularized semi-supervised extreme learning machine for EEG classification. *Int. J. Mach. Learn. Cyb.* 12, 903–916. doi: 10.1007/s13042-020-01209-0

Shen, Y. W., and Lin, Y. P. (2019). Challenge for affective brain-computer interfaces: non-stationary spatio-spectral EEG oscillations of emotional responses. *Front. Hum. Neurosci.* 13:366. doi: 10.3389/fnhum.2019.00366

Singh, A., Lal, S., and Guesgen, H. (2019). Small sample motor imagery classification using regularized riemannian features. *IEEE Access* 7, 46858–46869. doi: 10.1109/ACCESS.2019.2909058

Soekadar, S., Witkowski, M., Gómez, C., Opisso, E., Medina, J., Cortese, M., et al. (2016). Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia. *Sci. Robot* 1, 32–96. doi: 10.1126/scirobotics.aag3296

Sybeldon, M., Schmit, L., and Akcakaya, M. (2017). Transfer learning for SSVEP electroencephalography based brain–computer interfaces using learn++.NSE and mutual information. *Entropy* 19:41. doi: 10.3390/e19010041

Tang, X., Wang, T., Du, Y., and Dai, Y. (2019). Motor imagery EEG recognition with KNN-based smooth auto-encoder. *Artif. Intell. Med.* 101:101747. doi: 10.1016/j.artmed.2019.101747

Tkachenko, M., and Lauw, H. W. (2017). Comparative relation generative model. *IEEE Trans. Knowl. Data Eng.* 29, 771–783. doi: 10.1109/TKDE.2016.2640281

Vidal, J. (1977). Real-time detection of brain events in EEG. *Proc. IEEE* 65, 633–641. doi: 10.1109/PROC.1977.10542

Wang, H., and Li, X. (2016). Regularized filters for L1-norm-based common spatial patterns. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 201–211. doi: 10.1109/TNSRE.2015.2474141

Wang, J., Gu, Z., Yu, Z., and Li, Y. (2017). An online semi-supervised P300 speller based on extreme learning machine. *Neurocomputing* 269, 148–151. doi: 10.1016/j.neucom.2016.12.098

Wang, P., Lu, J., Zhang, B., and Tang, Z. (2015). "A review on transfer learning for brain-computer interface classification," in *Proc. 5th Int'l Conf. on Information Science and Technology*, (Changsha), 315–322. doi: 10.1109/icist.2015.7288989

Wei, C., Lin, Y., Wang, Y., Jung, T., Biggdely-Shamlo, N., and Lin, C. (2015). "Selective transfer learning for EEG-based drowsiness detection," in *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (Hongkong), 3229–3232. doi: 10.1109/SMC.2015.560

Wei, C., Lin, Y., Wang, Y., Lin, C., and Jung, T. (2018). A subject-transfer framework for obviating inter- and intra-subject variability in EEG-based drowsiness detection. *NeuroImage* 174, 407–419. doi: 10.1016/j.neuroimage.2018.03.032

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Wu, D. (2017a). "Active semi-supervised transfer learning (ASTL) for offline BCI calibration," in *Proceedings of the 2017 IEEE International Conference on Systems, Man and Cybernetics (SMC) (Banff: IEEE)*, (Banff, AB), 246–251. doi: 10.1109/SMC.2017.8122610

Wu, D. (2017b). Online and offline domain adaptation for reducing BCI calibration effort. *IEEE Trans. Hum. Mach. Syst.* 47, 550–563. doi: 10.1109/THMS.2016.2608931

Wu, D., Lawhern, V. J., Gordon, S., Lance, B. J., and Lin, C. (2016). Driver drowsiness estimation from EEG signals using online weighted adaptation regularization for regression (OwARR). *IEEE Trans. Fuzzy Syst.* 25, 1522–1535. doi: 10.1109/TFUZZ.2016.2633379

Xu, Y., Hua, J., Zhang, H., Hu, R., Huang, X., Liu, J., et al. (2019a). Improved transductive support vector machine for a small labelled set in motor imagery-based brain-computer interface. *Comput. Intel. Neurosc.* 2019:2087132. doi: 10.1155/2019/2087132

Xu, Y., Wei, Q., Zhang, H., Hu, R., Liu, J., Hua, J., et al. (2019b). Transfer learning based on regularized common spatial patterns using cosine similarities of spatial filters for motor-imagery BCI. *J. Circuit. Syst. Comp.* 28:1950123. doi: 10.1142/S0218126619501238

Yin, E., Zeyl, T., Saab, R., Chau, T., Hu, D., and Zhou, Z. (2015). A hybrid brain–computer interface based on the fusion of P300 and SSVEP scores. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 693–701. doi: 10.1109/TNSRE.2015.2403270

Yin, Z., Wang, Y., Liu, L., Zhang, W., and Zhang, J. (2017). Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. *Front. Neurorobot.* 11:19. doi: 10.3389/fnbot.2017.00019

Yong, L., and Yu, M. (2020). Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning. *Biomed. Signal Proces.* 62:102101. doi: 10.1016/j.bspc.2020.102101

Zanini, P., Congedo, M., Jutten, C., Said, S., and Berthoumieu, Y. (2018). Transfer learning: a Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 65, 1107–1116. doi: 10.1109/TBME.2017.2742541

Zhang, K., Xu, G., Zheng, X., Li, H., Zhang, S., Yu, Y., et al. (2020). Application of transfer learning in EEG decoding based on brain-computer interfaces: a review. *Sensors* 20:6321. doi: 10.3390/s20216321

Zhang, W., and Wu, D. (2020). Manifold embedded knowledge transfer for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1117–1127. doi: 10.1109/TNSRE.2020.2985996

Zhang, Y., Chen, J., Tan, J. H., Chen, Y., Li, D., Yang, L., et al. (2020). An investigation of deep learning models for EEG-based emotion recognition. *Front. Hum. Neurosci.* 14:622759. doi: 10.3389/fnins.2020.622759

Zhang, Y., Zhao, Q., Jin, J., Wang, X., and Cichocki, A. (2012). A novel BCI based on ERP components sensitive to configural processing of human faces. *J. Neural Eng.* 9:026018. doi: 10.1088/1741-2560/9/2/026018

Zhao, M., Chow, T. W. S., Zhang, Z., and Bing, L. (2015). Automatic image annotation via compact graph based semi-supervised learning. *Knowl. Based Syst.* 76, 148–165. doi: 10.1016/j.knosys.2014.12.014

Zhao, X., Zhao, J., Cai, W., and Wu, S. (2019). Transferring common spatial filters with semi-supervised learning for zero-training motor imagery brain-computer interface. *IEEE Access* 7, 58120–58130. doi: 10.1109/ACCESS.2019.2913154

Zheng, M., Yang, B., and Xie, Y. (2020). EEG classification across sessions and across subjects through transfer learning in motor imagery-based brain-machine interface system. *Med. Biol. Eng. Comput.* 58, 1515–1528. doi: 10.1007/s11517-020-02176-y

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zheng, X., Li, J., Ji, H., Duan, L., Li, M., Pang, Z., et al. (2020). Task transfer learning for EEG classification in motor imagery-based BCI system. *Comput. Math. Method. Med.* 2020, 1–11. doi: 10.1155/2020/6056383

Zhou, Y., She, Q., Ma, Y., Kong, W., and Zhang, Y. (2021). Transfer of semi-supervised broad learning system in electroencephalography signal classification. *Neural Comput. Appl.* 33, 10597–10613. doi: 10.1007/s00521-021-05793-2

Zhu, X., and Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synth. Lect. Artif. Intell. Mach. Learn.* 3, 1–130. doi: 10.2200/S00196ED1V01Y200906AIM006

Zou, Y., Zhao, X., Chu, Y., Zhao, Y., Xu, W., and Han, J. (2018). An inter-subject model to reduce the calibration time for motion imagination-based brain-computer interface. *Med. Biol. Eng. Comput.* 57, 939–952. doi: 10.1007/s11517-018-1917-x

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 10 September 2021
doi: 10.3389/fnins.2021.738268

Identification of Epileptic EEG Signals Through TSK Transfer Learning Fuzzy System

*Zhaoliang Zheng[1], Xuan Dong[2], Jian Yao[1], Leyuan Zhou[3], Yang Ding[3] and Aiguo Chen[1]**

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China, [2] Department of Tuberculosis and Respiratory Diseases, Wuhan Jinyintan Hospital, Wuhan, China, [3] Department of Radiotherapy, Affiliated Hospital, Jiangnan University, Wuxi, China

We propose a new model to identify epilepsy EEG signals. Some existing intelligent recognition technologies require that the training set and test set have the same distribution when recognizing EEG signals, some only consider reducing the marginal distribution distance of the data while ignoring the intra-class information of data, and some lack of interpretability. To address these deficiencies, we construct a TSK transfer learning fuzzy system (TSK-TL) based on the easy-to-interpret TSK fuzzy system the transfer learning method. The proposed model is interpretable. By using the information contained in the source domain and target domains more effectively, the requirements for data distribution are further relaxed. It realizes the identification of epilepsy EEG signals in data drift scene. The experimental results show that compared with the existing algorithms, TSK-TL has better performance in EEG recognition of epilepsy.

Keywords: epilepsy EEG signals, TSK fuzzy system, transfer learning, knowledge learning, interpretability

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Yanhui Zhang, Hebei University of Chinese Medicine, China
Min Shi, Fuzhou University of International Studies and Trade, China

****Correspondence:***
Aiguo Chen
agchen@jiangnan.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *08 July 2021*
Accepted: *16 August 2021*
Published: *10 September 2021*

Citation:
Zheng Z, Dong X, Yao J, Zhou L, Ding Y and Chen A (2021) Identification of Epileptic EEG Signals Through TSK Transfer Learning Fuzzy System. Front. Neurosci. 15:738268. doi: 10.3389/fnins.2021.738268

INTRODUCTION

Epilepsy is a disease caused by the sudden discharge of cerebral neurons. EEG technology (Suk et al., 2018) can monitor the changes of brain electrical signals, so we often use EEG intelligent recognition technology to detect epilepsy (Litt et al., 2001; Iasemidis et al., 2003; Dorai and Ponnambalam, 2010). Nowadays, many machines learn algorithms (Wang et al., 2015; Zhang et al., 2021) have been used to recognize epileptic signals, such as Decision Tree (Wang et al., 2015), nearest neighbor (KNN) (Iscan et al., 2011), Naive Bayes Algorithm (NB; Iscan et al., 2011), support vector machines (SVM; Yang et al., 2014), and fuzzy system (Aarabi et al., 2009; Rabbi and Fazel-Rezai, 2012; Deng et al., 2014a,b, 2018; Jiang et al., 2015). It has been proved that these algorithms can detect epilepsy faster and more accurately than doctors. However, as shown in **Figure 1**, only when the training set and test set obey the similar distribution can they show good classification performance. However, in most cases, as shown in **Figure 2**, the distribution characteristics of EEG data are not exactly the same. In order to make full use of their similar information, some researchers have proposed to use transfer learning algorithm, such as LMPROJ (Yang et al., 2014) and STL (Wang et al., 2018), applying the old knowledge we have gained to new fields. Different from traditional machine learning, which acquires knowledge from data and applies it to new problems, transfer learning focuses on transferring the learned knowledge to new

problems. Although the problem of different data distribution has been solved to a certain extent, these transfer learning algorithms only consider reducing the marginal distribution probability or conditional distribution probability (Deng et al., 2018) of data, without comprehensive balance, and these algorithms lack of interpretability.

To solve these problems, we propose a new method of EEG recognition based on transfer learning and a fuzzy system. The traditional method has a single model structure, so it cannot achieve good results in the face of complex scenes. Different from the traditional method, we pay attention to how to make full use of the previously marked data while ensuring the accuracy of the model on the new task. We not only minimize the marginal distribution or conditional distribution probability, but also combine them to minimize the joint probability distribution (Deng et al., 2018), and reach the best balance between marginal distribution and conditional distribution. In terms of ensuring interpretability, we use the TSK fuzzy system. Its IF-THEN rules can help us understand the rules of model operation more clearly. It has been widely used in data flow modeling, mining tasks, metacognitive learning, and multi-task learning. In order to realize this system, a TSK fuzzy system construction method based on transfer learning (TSK-TL) was developed. The experimental results show that compared with the existing algorithms, TSK-TL has better performance in EEG recognition of epilepsy.

Our contributions are mainly reflected on: (1) The introduction of transfer learning technology (Wang and Mahadevan, 2011; Quanz et al., 2012; Xiao and Guo, 2012), the proposed model in ensuring the accuracy of recognition At the same time, it has higher interpretability; (2) It has stronger robustness and can handle more complex data scenes; (3) It realizes the more accurate identification of epileptic EEG signals in data drift scenarios.

The rest of the manuscript is organized as follows. In the section "Backgrounds,"we briefly introduced the EEG data set, the classical TSK model and the related contents of transfer learning. In section "Identification of Epileptic EEG Signals Through TSK Transfer Learning fuzzysystem,"we first introduced the framework based on transfer learning, and then proposed the objective function of the TSK-TL. In section "Experimental Process and Result Analysis," we introduce the details of our experiment to test the performance of TSK-TL. The conclusion is given in the last section.

BACKGROUNDS

This section introduces the data sets and their processing methods used in the research, the classical TSK fuzzy system and the related content of transfer learning.

Epilepsy EEG Signal Dataset

The original epileptic EEG data set used in this study is divided into five groups, i.e., Group A to Group E, each group contains 100 single-channel signal segments, and the sampling rate of all samples is adjusted to 173.6 Hz. Among them, the data from healthy people are divided into groups A and B. The difference is that the eyes of group A are opened and group B is closed. The data of groups C, D, and E are obtained from volunteers with epilepsy in different states (Li, 2021). **Figure 3** shows the five groups of original epilepsy EEG signals. **Table 1** presents these five groups in detail.

The distribution law of EEG signals changes with time. Its amplitude is very small, and it is easy to be affected by other human biological currents such as ECG, EOG, and EMG. At the same time, it has strong randomness, and the noise in the signal is very complicated. Therefore, the experimental results obtained by using the original EEG signal directly are not ideal. According to previous work (Jiang et al., 2017; Tsujikawal et al., 2018), WPD, STFT, and KPCA are three typical feature extraction methods to process epileptic EEG signals (Blanco et al., 1997; Zhang et al., 2000; Srinivasan et al., 2005; Vivaldi and Bassi, 2006; Tzallas et al., 2009; Tang and Durand, 2012; Teng et al., 2017). **Figure 4** shows the sample of group A processed by the three feature extraction methods.

Classical TSK Fuzzy System

Because of its unique interpretability, fuzzy systems has been widely used in modeling and intelligent control. In addition, the output of TSK fuzzy system is more concise. The training process can be transformed into a linear regression problem or a quadratic programming problem, which makes the training process more efficient (Deng et al., 2012; Jiang et al., 2017).

The inference rules of TSK fuzzy system are usually defined as:

$$R^k: \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \cdots \wedge x_d \text{ is } A_d^k, \tag{1}$$
$$\text{Then } f^k(\mathrm{x}) = p_0^k + p_1^k x_1 + \cdots + p_d^k x_d, k = 1, \cdots, K$$

K is the number of fuzzy rules. Each rule is premised on the input vector $\mathrm{x} = [x_1, x_2, , x_d]^{\mathrm{T}}$ and maps the fuzzy set in the input space $\mathrm{A}^k \subset \mathrm{R}^d$ to the change single case represented by $f^k(\mathbf{x})$. A_i^k is the fuzzy subset of the ith dimension of the input vector x under the kth rule. $\wedge$ is a fuzzy conjunction operator. According to previous work (Jiang et al., 2017) the result of the TSK fuzzy model can be expressed as

$$y^0 = \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x}) f^k(\mathbf{x})}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x})} = \sum_{k=1}^{K} \tilde{\mu}^k(\mathbf{x}) f^k(\mathbf{x}) \tag{2.a}$$

where

$$\mu^k(\mathbf{x}) = \prod_{i=1}^{d} \mu_{A_i^k}(x_i) \tag{2.b}$$

and

$$\tilde{\mu}^k(\mathbf{x}) = \mu^k(\mathbf{x}) / \sum_{k'=1}^{K} \mu^{k'}(\mathbf{x}) \tag{2.c}$$

we use the Gaussian membership function, i.e.,

$$\mu_{A_i^k}(x_i) = \exp\left(\frac{-(x_i - c_i^k)^2}{2\delta_i^k}\right) \tag{2.d}$$

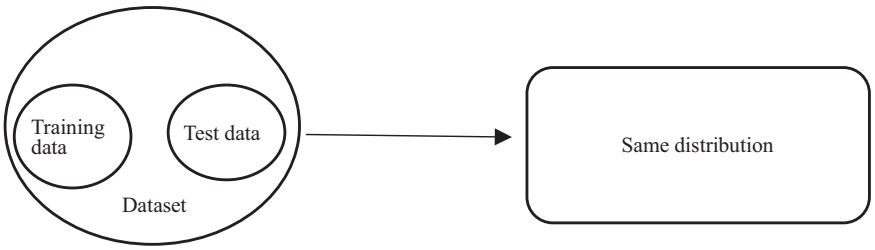


FIGURE 1 | Data distribution scenarios required by traditional methods.

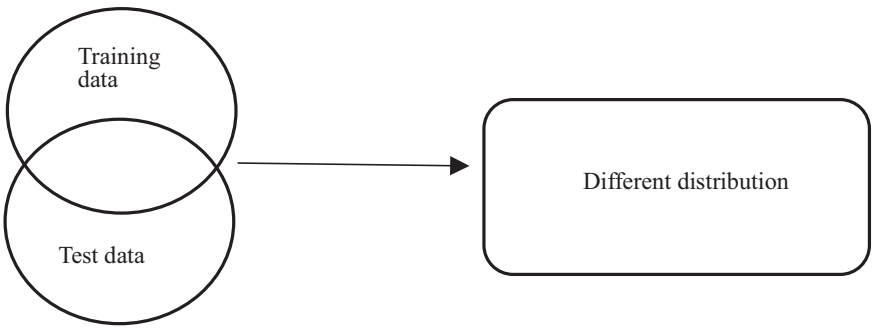


FIGURE 2 | Actual data distribution scenario.

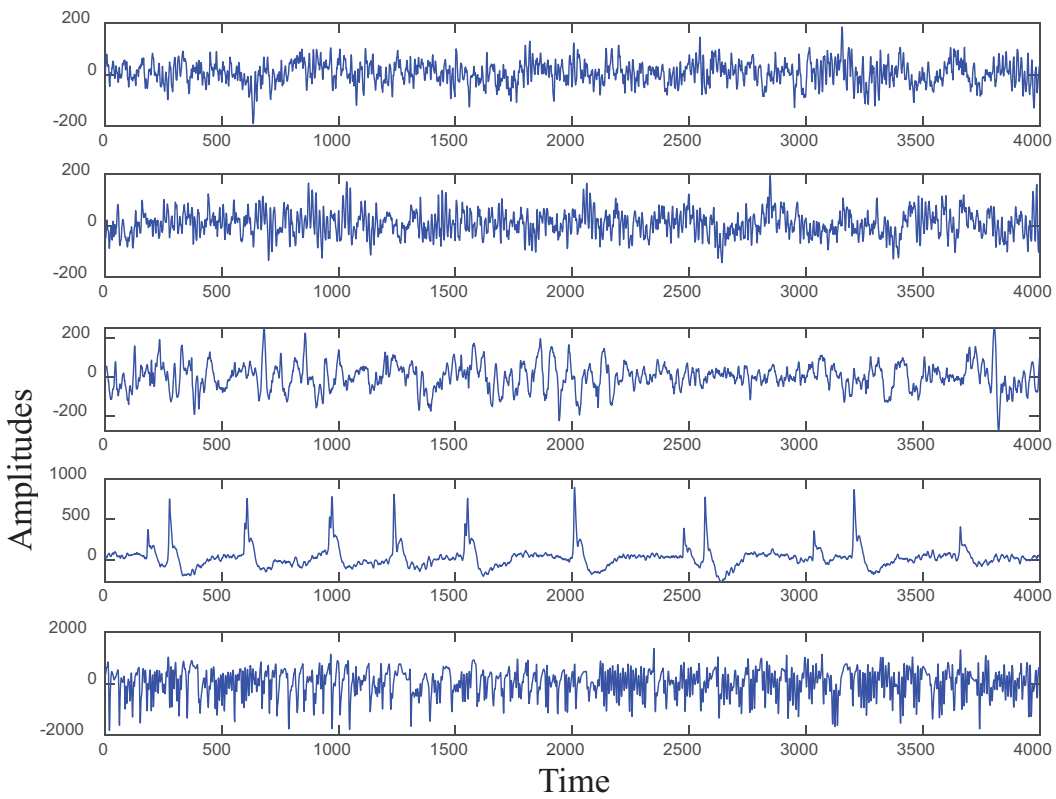


FIGURE 3 | Original epilepsy EEG signals.

TABLE 1 | Description of the dataset.

Volunteer	Group	Size of group	Description
Healthy volunteer	A	100	EEG signals in healthy volunteers with eyes open
	B	100	EEG signals in healthy volunteers with eyes closed
Volunteer with epilepsy	C	100	EEG signals of the hippocampal structure of the brain during the intermittent period of epilepsy
	D	100	EEG signals in epileptic areas of the brain during intermittent epilepsy
	E	100	EEG signals measured in epileptic patients during the seizure

as the fuzzy membership function. In the paper, we use the FCM algorithm to obtain c_i^k and δ_i^k. They can be estimated by the following expressions

$$c_i^k = \sum_{j=1}^{N} u_{jk} x_{ji} \Big/ \sum_{j=1}^{N} u_{jk} \tag{2.e}$$

$$\delta_i^k = h \cdot \sum_{j=1}^{N} u_{jk} (x_{ji} - c_i^k)^2 \Big/ \sum_{j=1}^{N} u_{jk} \tag{2.f}$$

u_{jk} is the fuzzy membership corresponding to the jth sample in the kth cluster. And h is the artificially adjustable scale parameter. After determining these antecedent parameters, let

$$\mathbf{x}_e = \left(1, \mathbf{x}^T\right)^T \tag{3.a}$$

$$\widetilde{\mathbf{x}}^k = \tilde{\mu}^k(\mathbf{x})\, \mathbf{x}_e \tag{3.b}$$

$$\mathbf{x}_g = \left(\left(\widetilde{\mathbf{x}}^1\right)^T, \left(\widetilde{\mathbf{x}}^2\right)^T, ..., \left(\widetilde{\mathbf{x}}^K\right)^T\right)^T \tag{3.c}$$

$$\mathbf{p}^k = \left(p_0^k, p_1^k, ..., p_d^k\right)^T \tag{3.d}$$

$$\mathbf{p}_g = \left(\left(\mathbf{p}^1\right)^T, \left(\mathbf{p}^2\right)^T, ..., \left(\mathbf{p}^K\right)^T\right)^T \tag{3.e}$$

According to the above transformation, Eq. 2a be converted to the following linear regression problem (Jiang et al., 2017).

$$y^o = \mathbf{p}_g^T \mathbf{x}_g \tag{3.f}$$

A well-performing algorithm is proposed (Jiang et al., 2017) to train the classic TSK-FS. The objective function of this algorithm is

$$\min_{\mathbf{P}_g} J_{TSK-FS}(\mathbf{P}_g) = \frac{1}{2}\mathbf{P}_g^{\mathbf{T}}\mathbf{P}_g + \frac{\lambda_1}{2}\left\|\mathbf{P}_g^{\mathrm{T}}\mathbf{X} - \mathbf{Y}\right\|^2 \tag{4}$$

where $\frac{1}{2}\mathbf{P}_g^{\mathbf{T}}\mathbf{P}_g$ is the regularization term, which can effectively promote the generalization ability of the TSK-FS; $\mathbf{P}_g$ is a consequent parameter; X is the matrix obtained by (5c); $_1$ is a regularization parameter. It can adjust the balance between model complexity and error tolerance; $[\mathrm{Y} = y_0, y_1, ..., y_n]$ is the label matrix.

In order to obtain the optimal $\mathbf{P}_g$, the derivative of $J_{TSK-FS}(\mathbf{P}_g)$ with respect to $\mathbf{P}_g$ can be set to 0, and then the optimal solution of $\mathbf{P}_g$ can be obtained as follows:

$$\mathbf{P}_g = \left(\mathbf{I} + \lambda_1 \cdot \mathbf{X}\mathbf{X}^{\mathbf{T}}\right)^{-1} \cdot (\lambda_1 \cdot \mathbf{X}\mathbf{Y}) \tag{5}$$

Through the optimal prior and posterior parameters, we can establish a classic TSK fuzzy system.

Preparatory Knowledge of Transfer Learning

There are three methods for transfer learning: data distribution adaptation, feature selection, and subspace learning (Shi et al., 2013; Zheng, 2021). The basic concept of data distribution adaptation is to make the probability distribution of the data of the source domain ($\mathcal{D}_s$) and the target domain ($\mathcal{D}_t$) the same or similar through some transformations. Feature selection method considers that the source domain and the target domain contain some common features, and their data distribution is similar. Then, the common features are extracted through the machine learning method, and the model can be built based on these features. The subspace learning method usually assumes that the data of the source domain and target domain

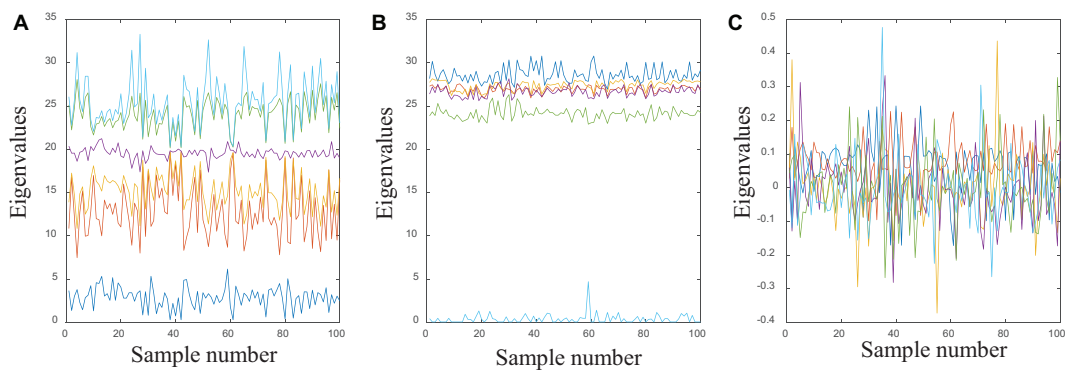


FIGURE 4 | (A) The samples of group A processed through WPD; **(B)** the samples of group A processed through STFT; **(C)** the samples of group A processed through KPCA.

will have similar distributions in the transformed subspace, and then learn through statistical feature transformation or manifold transformation.

The joint distribution adaptation (Wang et al., 2018) adopted in the paper belongs to the data distribution adaptive methods. Specifically, the core of joint distribution adaptation is to simultaneously minimize the marginal probability distribution and the conditional probability distribution of the two domains. The distance in machine learning has various forms. Here we use the MMD (Long et al., 2012) as a distance measurement. It can be calculated as follows:

$$MMD_H^2\left(D_s, D_t\right) = \left\| \frac{1}{n}\sum_{i=1}^{n} \emptyset\left(\mathbf{x}_i\right) - \frac{1}{m}\sum_{j=n+1}^{n+m} \emptyset\left(\mathbf{x}_j\right) \right\|^2 \quad (6)$$

where $\emptyset$ is a feature mapping, n is the number of samples in the source domain and m indicates the number of samples in the target domain.

IDENTIFICATION OF EPILEPTIC EEG SIGNALS THROUGH TSK TRANSFER LEARNING FUZZYSYSTEM

In this section, we will introduce in detail the transfer learning techniques we use. Combined with the analysis and research on the rules and parameter learning strategies of the classical TSK fuzzy system, a TSK-TL method for detecting epileptic signals is proposed.

Framework Based on Transfer Learning

The transfer learning strategy used in this study is divided into two parts: joint distribution adaptation (Zheng, 2021) and historical knowledge learning mechanism. As shown in **Figure 5**, the framework of epilepsy EEG signals recognition based on transfer learning theory is given. In order to make full use of the information of source domain and target domain, our work mainly includes three steps: (1) minimizing the marginal probability distribution, (2) minimizing the conditional probability distribution, and (3) further learning with historical knowledge.

(1) Minimize the marginal probability distribution

The marginal distributions of the source domain and target domain are represented by P_s and P_t, respectively. The paper builds a new model based on TSK-FS, $\mathbf{P}_g$can be taken as a projected vector, $\mathbf{x}_g$is a projected vector. And in order to make MMD a proper regularization for the classifier, we adopt the projected MMD (Long et al., 2013). Then the marginal probability distribution can be obtained by

$$D\left(P_s, P_t\right) = \left\| \frac{1}{n}\sum_{\mathrm{i}=1}^{n} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_{si} - \frac{1}{m}\sum_{\mathrm{j}=1}^{m} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_{tj} \right\|^2 \quad (7)$$

where $\mathbf{x}_{si}$ is the ith sample of the source domain, $\mathbf{x}_{tj}$ is the jth sample of the target domain, n and m indicate the number of samples in the source domain and target domain, respectively.

(2) Minimize the conditional probability distribution

Reducing the distance of conditional distribution is actually to achieve intra-class migration, but we don't know the label of the target domain. In the paper, we use some traditional classification algorithms (such as SVM) to obtain the pseudo-label of the target domain. At the same time, we assume that the calculated pseudo-class centroid may be located not far from the real class centroid (Long et al., 2012). Therefore, we can calculate the distance of conditional probability distribution by using both true label and pseudo label. The conditional probability distribution of source domain and target domain are represented by Q_s and Q_t, and then the conditional distribution can be calculated as follows:

$$D^{(c)}(Q_s, Q_t) = \left\| \frac{1}{n^{(c)}} \sum_{\mathbf{X}_i \in \mathcal{D}_s^{(c)}} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_i - \frac{1}{m^{(c)}} \sum_{\mathbf{X}_j \in \mathcal{D}_t^{(c)}} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_j \right\|^2 \quad (8)$$

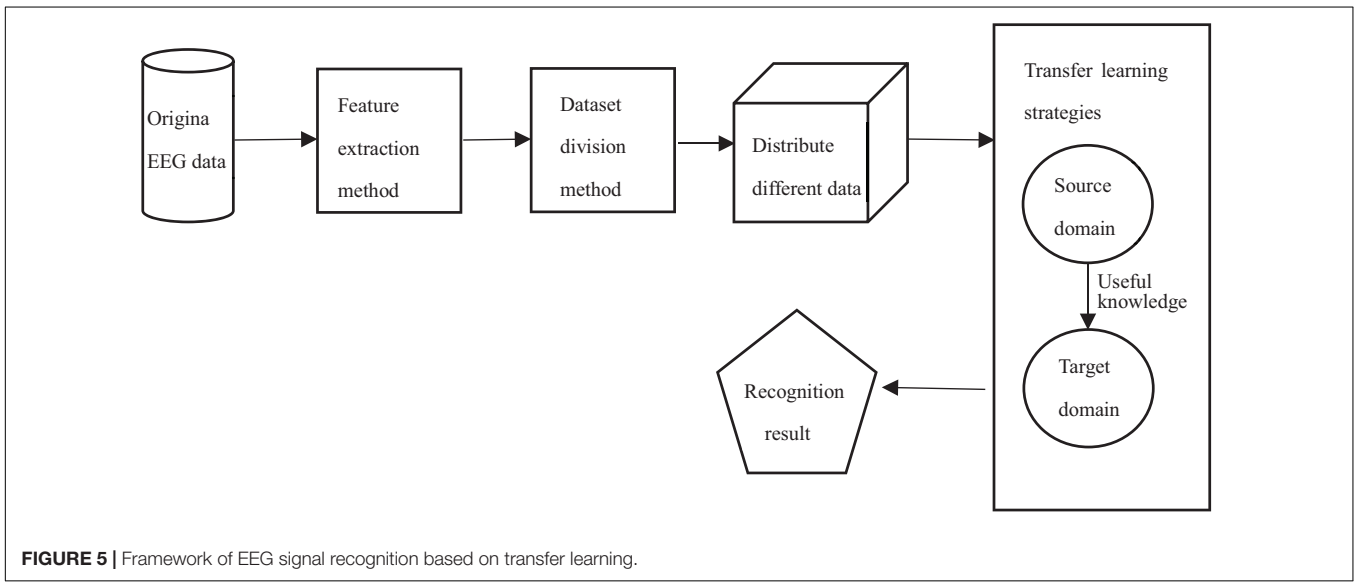


FIGURE 5 | Framework of EEG signal recognition based on transfer learning.

where $c \in \{0, 1, ..., C\}$ is the category tag. $\mathcal{D}_s^{(c)}$ is a set of the data belonging to class c in the source domain and $n^{(c)} = \left|\mathcal{D}_s^{(c)}\right|$. Correspondingly, $\mathcal{D}_t^{(c)}$ is a set of the data belonging to class c in the target domain and $m^{(c)} = \left|\mathcal{D}_t^{(c)}\right|$.

By integrating Equations 9 and 10, the joint probability distribution distance can be calculated as follows:

$$D(J_s, J_t) = D(P_s, P_t) + \sum_{c=1}^{C} D^{(c)}(Q_s, Q_t) \tag{9}$$

(3) Combine the historical knowledge to further learn

The parameter $\mathbf{P}_{g_0}$ obtained through classic TSK-FS is used to further guide the learning, and then the complete tranfer learning item is

$$D(\mathcal{D}_s, \mathcal{D}_t) = D(J_s, J_t) + \lambda_3 D\left(\mathbf{P}_{g_0}, \mathbf{P}_g\right) \tag{10}$$

where

$$D\left(\mathbf{P}_{g_0}, \mathbf{P}_g\right) = \|\mathbf{P}_{g_0} - \mathbf{P}_g\|^2 \tag{11}$$

The Objective Function of TSK-TL

We design the objective function of TSK-TL as

$$\min_{\mathbf{P}_g} J_{TSK-TL}\left(\mathbf{P}_g\right) = g + \lambda_2 D(J_s, J_t) + \lambda_3 D(\mathbf{P}_{g_0}, \mathbf{P}_g) \tag{12.a}$$

where

$$g = \frac{1}{2}\mathbf{P}_g^{\mathrm{T}}\mathbf{P}_g + \frac{\lambda_1}{2}\|\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_s - \mathbf{Y}\|^2 \tag{12.b}$$

$$\begin{aligned} D(J_s, J_t) &= D(P_s, P_t) + \sum_{c=1}^{C} D(Q_s, Q_t) \\ &= \left\| \frac{1}{n}\sum_{\mathrm{i}=1}^{n} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_{si} - \frac{1}{m}\sum_{\mathrm{j}=1}^{m} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_{tj} \right\|^2 \\ &+ \left\| \frac{1}{n^{(c)}} \sum_{\mathbf{X}_i \in D_s^{(c)}} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_i - \frac{1}{m^{(c)}} \sum_{\mathbf{X}_j \in D_t^{(c)}} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_j \right\|^2 \end{aligned} \tag{12.c}$$

$$D\left(\mathbf{P}_{g_0}, \mathbf{P}_g\right) = \|\mathbf{P}_{g_0} - \mathbf{P}_g\|^2 \tag{12.d}$$

where $\mathbf{P}_g$ is the expected projection of TSK-TL, $\mathbf{P}_{g_0}$ is the consequent parameter of the classical TSK model. $\mathbf{X}_s$ is a data matrix from the source domain. $\mathbf{Y} = [y_0, y_1, ..., y_n]$ is the label matrix, if the ith sample belongs to healthy volunteers, then y_i is 1, otherwise y_i is -1. And $_1$, $_2$, $_3$ are the regularization parameters.

Then we further explain the above formula as follows:

(1) Equation 12b is the training model of the classic TSK fuzzy system, so the TSK-TL we proposed inherits all its advantages.
(2) When experimenting with the classical TSK fuzzy system, the result is poor because of the distribution difference of data. By Equation 12c, the joint probability distribution distance is minimized to optimize the experimental results.
(3) In Equation 12d, $\mathbf{P}_g$ is further optimized by measuring the distance between $\mathbf{P}_{g_0}$ and $\mathbf{P}_g$. If $_3$ is infinite and the term is optimal, then $\mathbf{P}_{g_0}$ and $\mathbf{P}_g$ are equal.
(4) The regularization parameters $_1$ 0, $_2$ 0, and $_3$ 0 are used to control the balance between different terms. We use the grid search method to determine their values.

Solution of TSK-TL

In Equation 12c, the first term can be converted as follows:

$$\begin{aligned} &\left\| \frac{1}{n}\sum_{\mathrm{i}=1}^{n} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_{si} - \frac{1}{m}\sum_{\mathrm{j}=1}^{m} \mathbf{P}_g^{\mathrm{T}}\mathbf{x}_{tj} \right\|^2 \\ &= \left\| \frac{1}{n}\mathbf{P}_g^{\mathrm{T}}[\mathbf{x}_{s1}\ \mathbf{x}_{s2}\ \cdots\ \mathbf{x}_{sn}]_{1*n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n*1} \right. \\ &\qquad \left. - \frac{1}{m}\mathbf{P}_g^{\mathrm{T}}[\mathbf{x}_{t1}\ \mathbf{x}_{t2} \cdots \mathbf{x}_{tn}]_{1*m} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{m*1} \right\|^2 \\ &= \mathrm{tr}\left(\frac{1}{n^2}\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_s\mathbf{1}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_s\mathbf{1}\right)^{\mathrm{T}} + \frac{1}{m^2}\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_t\mathbf{1}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_t\mathbf{1}\right)^{\mathrm{T}} \right. \\ &\qquad \left. - \frac{1}{nm}\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_s\mathbf{1}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_s\mathbf{1}\right)^{\mathrm{T}} - \frac{1}{nm}\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_t\mathbf{1}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_t\mathbf{1}\right)^{\mathrm{T}} \right) \\ &\qquad + \mathrm{tr}\left[\mathbf{P}_g^{\mathrm{T}} \left(\frac{1}{n^2}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{X}_s^{\mathrm{T}}\mathbf{X}_s + \frac{1}{m^2}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{X}_t^{\mathrm{T}}\mathbf{X}_t \right. \right. \\ &\qquad \left. \left. - \frac{1}{nm}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{X}_s^{\mathrm{T}}\mathbf{X}_t - \frac{1}{nm}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{X}_t^{\mathrm{T}}\mathbf{X}_s \right) \mathbf{P}_g \right] \\ &= \mathrm{tr}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}\mathbf{M}_0\mathbf{X}^{\mathrm{T}}\mathbf{P}_g\right) \end{aligned}$$

Similarly, the second term can be converted as:

$$\mathrm{tr}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}\mathbf{M}_c\mathbf{X}^{\mathrm{T}}\mathbf{P}_g\right)$$

where $\mathbf{X}$ is a matrix composed of the source domain and target domain data. $\mathbf{M}_c$ is MMD matrix computed as:

$$(\mathbf{M}_c)_{ij} = \begin{cases} \frac{1}{n^{(c)}n^{(c)}}, \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{m^{(c)}m^{(c)}}, \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_t^{(c)} \\ \frac{-1}{n^{(c)}m^{(c)}}, \begin{cases} \mathbf{x}_i \in \mathcal{D}_s^{(c)}, \mathbf{x}_j \in \mathcal{D}_t^{(c)} \\ \mathbf{x}_j \in \mathcal{D}_s^{(c)}, \mathbf{x}_i \in \mathcal{D}_t^{(c)} \end{cases} \\ 0, \text{ otherwise} \end{cases} \tag{13}$$

So

$$D(J_s, J_t) = \mathrm{tr}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}\mathbf{M}\mathbf{X}^{\mathrm{T}}\mathbf{P}_g\right)\ \mathbf{M} = \sum_{c=0}^{C} \mathbf{M}_c$$

Then,

$$\min_{\mathbf{P}_g} J_{TSK-TL}\left(\mathbf{P}_g\right) = \frac{1}{2}\mathbf{P}_g^{\mathrm{T}}\mathbf{P}_g + \frac{\lambda_1}{2}\left\|\mathbf{P}_g^{\mathrm{T}}\mathbf{X}_s - \mathbf{Y}\right\|^2 + \lambda_2 \mathrm{tr}\left(\mathbf{P}_g^{\mathrm{T}}\mathbf{X}\mathbf{M}\mathbf{X}^{\mathrm{T}}\mathbf{P}_g\right) + \lambda_3\left\|\mathbf{P}_{g_0} - \mathbf{P}_g\right\|^2$$

By setting the derivative of J_{TSK-TL} with respect to $\mathbf{P}_g$ to zero, i.e., $\frac{\partial J_{TSK-TL}}{\mathbf{P}_g} = 0$, we can get the optimal solution of $\mathbf{P}_g$ as follows:

$$\mathbf{P}_g = \left(_2\left(\mathbf{X}\mathbf{M} + \mathbf{X}\mathbf{M}^{\mathrm{T}}\right)\mathbf{X}^{\mathrm{T}} + \lambda_1\mathbf{X}_s\mathbf{X}_s^{\mathrm{T}} + \lambda_3\mathbf{I}\right)^{-1}\left(\lambda_1\mathbf{X}_s\mathbf{Y}^{\mathrm{T}} + \lambda_3\mathbf{P}_{g_0}\right) \tag{14}$$

By (7), (15), and (16), we can obtain the optimal posterior parameter $\mathbf{P}_g$. Then based on $\mathbf{P}_g$, the final decision function can be obtained as follows

$$f\left(\mathbf{x}_{gi}\right) = \mathrm{sign}(\mathbf{P}_g\mathbf{x}_{gi}) \tag{15}$$

The details of the proposed TSK-TL algorithm are as follows:

Algorithm of TSK-TL.

Initialization: Set the number of fuzzy rules K, the regularization parameters λ_1, λ_2, and λ_3.

Stage 1: Construction of datasets for linear regression

Step 1: Through classical FCM or other partition techniques to divide the input space of training data to determine the premise of TSK-FS.

Step 2: Construct the new training dataset and test dataset through (3a)–(3c).

Stage 2: Computation of historical knowledge parameter $\mathbf{P}_{g_0}$

Step 3: Obtain the $\mathbf{P}_{g_0}$ by (5).

Stage 3: Obtain the parameter $\mathbf{P}_g$ ***of TSK-TL***

Step 4: Compute the MMD matrix by (13).

Step 5: Compute the consequent parameter $\mathbf{P}_g$ of TSK-TL through (14).

Step 6: Generate the TSK-TL by (15).

EXPERIMENTAL PROCESS AND RESULTS ANALYSIS

In this section, the proposed TSK-TL method is evaluated by classifying EEG signals of epilepsy patients and healthy people. In addition, a comparative study of five traditional machine learning algorithms and two transfer learning algorithms is carried out. Details of the experiments are as follows.

Experimental Setup

In this paper, we use three classical feature extraction methods, namely WPD, STFT, and KPCA, to obtain EEG datasets from three different views and perform experiments on them, respectively. In each experiment, we set up 10 experimental datasets, and every dataset is composed of part of the data from two or three different groups in this view. The details of these datasets are shown in **Table 2**. The structure of this dataset ensures that there is no or only a part of the data comimg from the same group. In short, there is no or only a part of the data in the source domain and the test domain has the same distribution.

TABLE 2 | Structure of experimental datasets.

Datasets	Source domain	Target domain
D1	Group A 50,Group E 50	Group A 50,Group C 50
D2	Group A 50,Group E 50	Group A 50,Group D 50
D3	Group B 50,Group E 50	Group B 50,Group C 50
D4	Group B 50,Group E 50	Group B 50,Group D 50
D5	Group A 50,Group E 50	Group B 50,Group C 50
D6	Group A 50,Group E 50	Group B 50,Group D 50
D7	Group B 50,Group E 50	Group A 50,Group D 50
D8	Group B 50,Group E 50	Group A 50,Group C 50
D9	Group A 50,Group B 50, Group E 50	Group A 50,Group B 50, Group C 50
D10	Group A 50,Group B 50, Group E 50	Group A 50,Group B 50, Group D 50

TABLE 3 | Experimental setup.

Methods for comparison	Classical recognition methods	Transfer learning recognition methods
	1. KNN 2. NB 3. SVM 4. C4.5 5. TSK-FS	1. LMPROJ 2. STL **3. TSK-TL**
Data label	1. EEG signals from healthy volunteers are labeled "1" 2. EEG signals from volunteers with epilepsy are labeled "−1"	
Performance evaluation	1. Accuracy: The ratio of the number of test data with correct prediction results to the total number of test data 2. Friedman test and Holm's post hoc test	
Method-specific settings	1. For TSK-FS, the number of fuzzy rules is taken the set {5, 10, 15, 20, 25, 30} and the regularization parameter λ_1 from the set $\left\{10^{-5}, 10^{-4}, ..., 10^4, 10^5\right\}$ 2. For the proposed TSK-TL, the number of fuzzy rules is taken from the set {5, 10, 15, 20, 25, 30}, and the regularization parameters λ_1, λ_2 and λ_3 from the set $\left\{10^{-5}, 10^{-4}, \cdots, 10^4, 10^5\right\}$ **Note:** Each parameter is determined by the grid search	

During the experiment, the experiment of each experimental dataset is repeated 10 times. And take the average result of 10 times to evaluate the performance. The optimal hyperparameters of each experimental model are determined by the grid search. All the algorithms are implemented using MATLAB on a computer with Intel i5-4590 3.3 GHz CPU, 12 GB of RAM. The details of the experimental setup are shown in **Table 3**.

Recognition Performance

In our experiment, five traditional machine learning algorithms and two transfer learning algorithms, i.e., LMPROJ and STL are used for comparative experiments. The results are shown in **Tables 4–6**. By observing these results, three conclusions can be drawn as follows.

(1) The proposed TSK-TL method not only reduces the difference of data distribution, but also optimizes the formation of new knowledge parameters through the prior knowledge parameters obtained from the classical TSK-FS model. The results show that our method can improve the accuracy of EEG signal recognition.

TABLE 4 | Experimental results of eight classifiers in the WPD view.

EEG datasets		Classifiers							
		KNN	NB	SVM	C4.5	TSK-FS	STL	LMPR-OJ	TSK-TL
D1	Mean (SD)	0.8771 (0.0275)	0.5636 (0.0180)	0.9086 (0.0135)	0.8845 (0.0254)	0.9157 (0.0079)	0.8529 (0.0160)	0.9286 (0.0157)	0.9586 (0.0168)
D2	Mean (SD)	0.8857 (0.0223)	0.6127 (0.0119)	0.9300 (0.0153)	0.8818 (0.0218)	0.9329 (0.0170)	0.8843 (0.0321)	0.9357 (0.0127)	0.9571 (0.0111)
D3	Mean (SD)	0.9623 (0.0113)	0.8373 (0.0385)	0.9571 (0.0160)	0.9645 (0.0104)	0.9614 (0.0069)	0.9789 (0.0038)	0.9773 (0.0049)	0.9971 (0.0049)
D4	Mean (SD)	0.9819 (0.0049)	0.8345 (0.0550)	0.9600 (0.0100)	0.9655 (0.0082)	0.9586 (0.0090)	0.9717 (0.0049)	0.9835 (0.0071)	0.9957 (0.0053)
D5	Mean (SD)	0.7529 (0.0472)	0.2655 (0.0151)	0.5900 (0.0183)	0.7936 (0.0566)	0.7271 (0.0287)	0.8629 (0.0522)	0.6829 (0.0801)	0.9543 (0.0199)
D6	Mean (SD)	0.7286 (0.0313)	0.3164 (0.0234)	0.5943 (0.0140)	0.7745 (0.0591)	0.7571 (0.0298)	0.8457 (0.0544)	0.6000 (0.0870)	0.9329 (0.0359)
D7	Mean (SD)	0.9157 (0.0230)	0.7927 (0.0447)	0.7914 (0.0313)	0.7727 (0.0174)	0.9043 (0.0190)	0.9171 (0.0298)	0.8900 (0.0258)	0.9614 (0.0234)
D8	Mean (SD)	0.8914 (0.0267)	0.8309 (0.0411)	0.7843 (0.0113)	0.8136 (0.0344)	0.8986 (0.0302)	0.9297 (0.0168)	0.8757 (0.0172)	0.9571 (0.0256)
D9	Mean (SD)	0.9143 (0.0254)	0.7152 (0.0233)	0.7439 (0.0101)	0.8994 (0.0212)	0.9217 (0.0112)	0.9367 (0.0159)	0.9286 (0.0171)	0.9714 (0.0120)
D10	Mean (SD)	0.9057 (0.0190)	0.7382 (0.0161)	0.7305 (0.0127)	0.9006 (0.0232)	0.9189 (0.0071)	0.9133 (0.0154)	0.9190 (0.0141)	0.9657 (0.0071)

TABLE 5 | Experimental results of eight classifiers in the STFT view.

EEG datasets		Classifiers							
		KNN	NB	SVM	C4.5	TSK-FS	STL	LMPROJ	TSK-TL
D1	Mean (SD)	0.6390 (0.0265)	0.5682 (0.0227)	0.9729 (0.0111)	0.8336 (0.0150)	0.8343 (0.1409)	0.8871 (0.0138)	0.8828 (0.0138)	0.9800 (0.0100)
D2	Mean (SD)	0.6575 (0.0341)	0.6309 (0.0187)	0.9429 (0.0137)	0.8373 (0.0142)	0.8368 (0.0548)	0.9014 (0.0107)	0.8771 (0.0143)	0.9857 (0.0079)
D3	Mean (SD)	0.5823 (0.0322)	0.5964 (0.0211)	0.9243 (0.0113)	0.7427 (0.1144)	0.9543 (0.0190)	0.8997 (0.0117)	0.9157 (0.0121)	0.9976 (0.0032)
D4	Mean (SD)	0.6060 (0.0481)	0.6409 (0.0489)	0.9647 (0.0053)	0.6927 (0.0827)	0.9171 (0.0496)	0.9167 (0.0026)	0.9529 (0.0049)	0.9986 (0.0053)
D5	Mean (SD)	0.62300 (0)	0.5473 (0.0233)	0.6029 (0.0150)	0.6727 (0.0236)	0.5943 (0.0812)	0.8014 (0.0324)	0.6314 (0.0564)	0.8583 (0.0676)
D6	Mean (SD)	0.5147 (0.0379)	0.4209 (0.0130)	0.5829 (0.0236)	0.6891 (0.0329)	0.5786 (0.1289)	0.7335 (0.0191)	0.6134 (0.0437)	0.8371 (0.0399)
D7	Mean (SD)	0.6015 (0.0452)	0.5909 (0.0212)	0.8400 (0.0163)	0.7127 (0.0163)	0.9071 (0.0160)	0.9046 (0.0107)	0.9171 (0.0160)	0.9786 (0.0079)
D8	Mean (SD)	0.6390 (0.0335)	0.6145 (0.0273)	0.8629 (0.0221)	0.6455 (0.0281)	0.9171 (0.0206)	0.9217 (0.0121)	0.9083 (0.0163)	0.9657 (0.0111)
D9	Mean (SD)	0.6837 (0.0136)	0.6739 (0.0020)	0.8990 (0.0186)	0.8133 (0.0149)	0.8805 (0.0214)	0.9049 (0.0076)	0.9129 (0.0065)	0.9857 (0.0067)
D10	Mean (SD)	0.7130 (0.0183)	0.7048 (0.0052)	0.9162 (0.0076)	0.7818 (0.0085)	0.8967 (0.0144)	0.9123 (0.0233)	0.9076 (0.0081)	0.9771 (0.0115)

TABLE 6 | Experimental results of eight classifiers in the KPCA view.

EEG datasets		Classifiers							
		KNN	NB	SVM	C4.5	TSK-FS	STL	LMPROJ	TSK-TL
D1	Mean (SD)	0.8813 (0.0336)	0.7760 (0.0219)	0.8167 (0.0462)	0.8955 (0.0121)	0.8700 (0.0394)	0.8414 (0.0308)	0.7781 (0.0164)	0.9120 (0.0311)
D2	Mean (SD)	0.9050 (0.0225)	0.8940 (0.0477)	0.9067 (0.0321)	0.9789 (0.0083)	0.9557 (0.0336)	0.9214 (0.0090)	0.8940 (0.0351)	0.9820 (0.0130)
D3	Mean (SD)	0.9295 (0.0206)	0.9120 (0.0152)	0.8500 (0.0361)	0.8382 (0.0325)	0.9360 (0.0152)	0.9200 (0.0100)	0.8120 (0.0130)	0.9500 (0.0187)
D4	Mean (SD)	0.9450 (0.0176)	0.9260 (0.0114)	0.8367 (0.0252)	0.9591 (0.0138)	0.9680 (0.0045)	0.9343 (0.0127)	0.8127 (0.0158)	0.9820 (0.0084)
D5	Mean (SD)	0.7034 (0.0526)	0.5560 (0.0195)	0.6067 (0.0115)	0.5945 (0.0543)	0.7240 (0.0658)	0.8217 (0.0492)	0.7020 (0.0327)	0.8289 (0.0430)
D6	Mean (SD)	0.7045 (0.0331)	0.6137 (0.0122)	0.6833 (0.0451)	0.7700 (0.0300)	0.7380 (0.0311)	0.8700 (0.0336)	0.7423 (0.0268)	0.8743 (0.0217)
D7	Mean (SD)	0.8827 (0.0239)	0.8109 (0.0138)	0.8700 (0.0400)	0.9218 (0.0271)	0.8943 (0.0292)	0.9186 (0.0135)	0.8343 (0.0151)	0.9420 (0.0228)
D8	Mean (SD)	0.8945 (0.0331)	0.8064 (0.0234)	0.8467 (0.0058)	0.7682 (0.0282)	0.8780 (0.0277)	0.8471 (0.0275)	0.8000 (0.0200)	0.9200 (0.0158)
D9	Mean (SD)	0.9063 (0.0158)	0.7358 (0.0414)	0.6689 (0.0038)	0.8345 (0.0333)	0.8687 (0.0197)	0.8771 (0.0148)	0.8667 (0.0054)	0.9280 (0.0179)
D10	Mean (SD)	0.9533 (0.0169)	0.7655 (0.0409)	0.6667 (0.0017)	0.9810 (0.0054)	0.9520 (0.0073)	0.9210 (0.0178)	0.9371 (0.0143)	0.9827 (0.0060)

(2) Compared with the two transfer learning classifiers LMPROJ and STL, TSK-TL is superior to LMPROJ in recognition of epileptic EEG signals, and its performance is better than or at least comparable to the STL method. STL method uses affinity within the class to transfer knowledge within the class, and then to learn. The LMPROJ method only reduces the marginal distribution of data based on the SVM algorithm.

(3) The TSK-TL method we proposed has a good recognition effect in the three views, and the average recognition accuracy in the WPD view is the highest, which points out the direction for us to select an appropriate feature extraction method in practical applications. At the same time, it can be observed that when processing the D5 and D6 data sets, the recognition accuracy is greatly improved compared with other methods, which further verifies that our proposed method has a better effect in the face of data with large differences in data distribution.

Statistical Analysis

We evaluated the experimental results from a statistical point of view through the Friedman test (Demšar, 2006; Garcia and Herrera, 2008) and the Holm's *post hoc* test (Demšar, 2006; Garcia and Herrera, 2008). Friedman test is used to calculate the average

TABLE 7 | Comparison of average performance of TSK-TI and other seven methods using Friedman test (A = 0.05).

Friedman Test for TSK-TL			
Algorithms	**Friedman Rank**	***p*-Value**	**H_0**
KPCA view			
KNN	3.8	0.000000006179	Rejected
NB	7.05		
SVM	6.4		
C4.5	4.2		
TSK-FS	3.4		
STL	3.8		
LMPROJ	6.35		
TSK-TL	1		
STFT view			
KNN	7	0.00000001016	Rejected
NB	7.8		
SVM	3.6		
C4.5	5.3		
TSK-FS	4.6		
STL	3.2		
LMPROJ	3.5		
TSK-TL	1		
WPD view			
KNN	4.5	0.000000004808	Rejected
NB	7.5		
SVM	6.5		
C4.5	5.4		
TSK-FS	4.2		
STL	3.3		
LMPROJ	3.6		
TSK-TL	1		

TABLE 8 | Holm's *post hoc* comparison of the results of Friedman procedure of TSK-TL and other seven methods

Holm's *post hoc* comparison for TSK-TL					
i	**Algorithm**	**$Z = (R_0-R_t)/SE$**	***p***	**Holm = /i**	**H_0**
KPCA view					
7	NB	5.522869	0	0.007142	Rejected
6	SVM	4.929503	0.000001	0.008333	Rejected
5	LMPROJ	4.883859	0.000001	0.01	Rejected
4	C4.5	2.921187	0.003487	0.0125	Rejected
3	KNN	2.556039	0.010587	0.016667	Rejected
2	STL	2.556039	0.010587	0.025	Rejected
1	TSK-FS	2.19089	0.02846	0.05	Rejected
STFT view					
7	NB	6.207522	0	0.007142	Rejected
6	KNN	5.477226	0	0.008333	Rejected
5	C4.5	3.925345	0.000087	0.01	Rejected
4	TSK-FS	3.286335	0.001015	0.0125	Rejected
3	SVM	2.373464	0.017622	0.016667	Rejected
2	LMPROJ	2.282177	0.022479	0.025	Rejected
1	STL	2.008316	0.04461	0.05	Rejected
WPD view					
7	NB	5.933661	0	0.007142	Rejected
6	SVM	5.02079	0.000001	0.008333	Rejected
5	C4.5	4.016632	0.000059	0.01	Rejected
4	KNN	3.195048	0.001398	0.0125	Rejected
3	TSK-FS	2.921187	0.003487	0.016667	Rejected
2	LMPROJ	2.373464	0.017622	0.025	Rejected
1	STL	2.099603	0.035764	0.05	Rejected

ranking of the compared methods and to determine whether the observed difference are statistically significant. We set the significance level of the test to 0.05. If the *p*-value is less than 0.05, the null hypothesis H0 is rejected and we can confirm that there are significant difference. The Holm's *post hoc* test is then performed to evaluate the statistical differences between the control (i.e., the method that achieves the best Friedman rank) and other methods. The results of the Friedman test are shown in **Table 7**, and the results of the Holm's *post hoc* test are shown in **Table 8**.

As shown in **Table 7**, the Friedman test results reveal that the TSK-TL method performs better than the other seven classification methods in classification accuracy. And the results

TABLE 9 | TSK-TL Model with five rules trained on the dataset D2 for the KPCA view.

Fuzzy rules base

TSK Fuzzy Rule R^k: IFx_1 *is* $A_1^k\left(c_1^k, \delta_1^k\right) \wedge x_2$ *is* $A_2^k\left(c_2^k, \delta_2^k\right) \wedge \cdots \wedge x_d$ *is* $A_d^k\left(c_d^k, \delta_d^k\right)$, *Then* $f_x(x) = p_k^0 + p_k^1 x_1 + \cdots p_k^d x_d$

No. of rules	Antecedent parameters	Consequent parameters
K	$c^k = (c_1^k, ..., c_d^k)^T, \delta^k = (\delta_1^k, ...\delta_d^k)^T$	$P_k = (P_k^0, P_k^1, ..., P_k^d)^T$
1	c^1 = [−0.1589,0.0712,0.0076,0.0517,−0.000356,0.0097] δ^1 [0.0029,0.0037,0.0052,0.0063,0.0039,0.0032]	p_1 = [−2.7245,−0.5118,0.9243,−0.295,−0.0378,−0.58,−1.1201]
2	c^2 = [0.0582,−0.0035,−0.0439,0.1377,−0.0389,−0.0297] δ^2 [0.0033,0.0021,0.0025,0.0056,0.0017,0.0042]	p_2 = [10.9842,−8.5028,−0.9881,4.7443,−1.8692,3.2596,−3.0533]
3	c^3 = [0.0035,0.1507,0.0516,−0.1698,−0.1519,0.0789] δ^3 [0.0022,0.0045,0.0062,0.0082,0.0051,0.0057]	p_3 = [2.045,3.995,−4.3315,−1.4117,−0.9264,−0.0619,0.4859]
4	c^4 = [0.1562,−0.0587,0.139,0.0092,0.1228,−0.1636] δ^4 [0.0038,0.0035,0.005,0.0054,0.0033,0.0072]	p_4 = [−3.8877,−2.2364,−0.7276,−1.1165,0.1062,−1.3669,0.9594]
5	c^5 = [−0.0667,−0.1566,−0.1623,−0.0287,0.0347,0.1368] *delta*^{5} [0.0026,0.0032,0.0047,0.0031,0.0036,0.0071]	p_5 = [2.0489,4.0366,−4.2784,−1.4602,−1.0065,0.0086,0.9774]

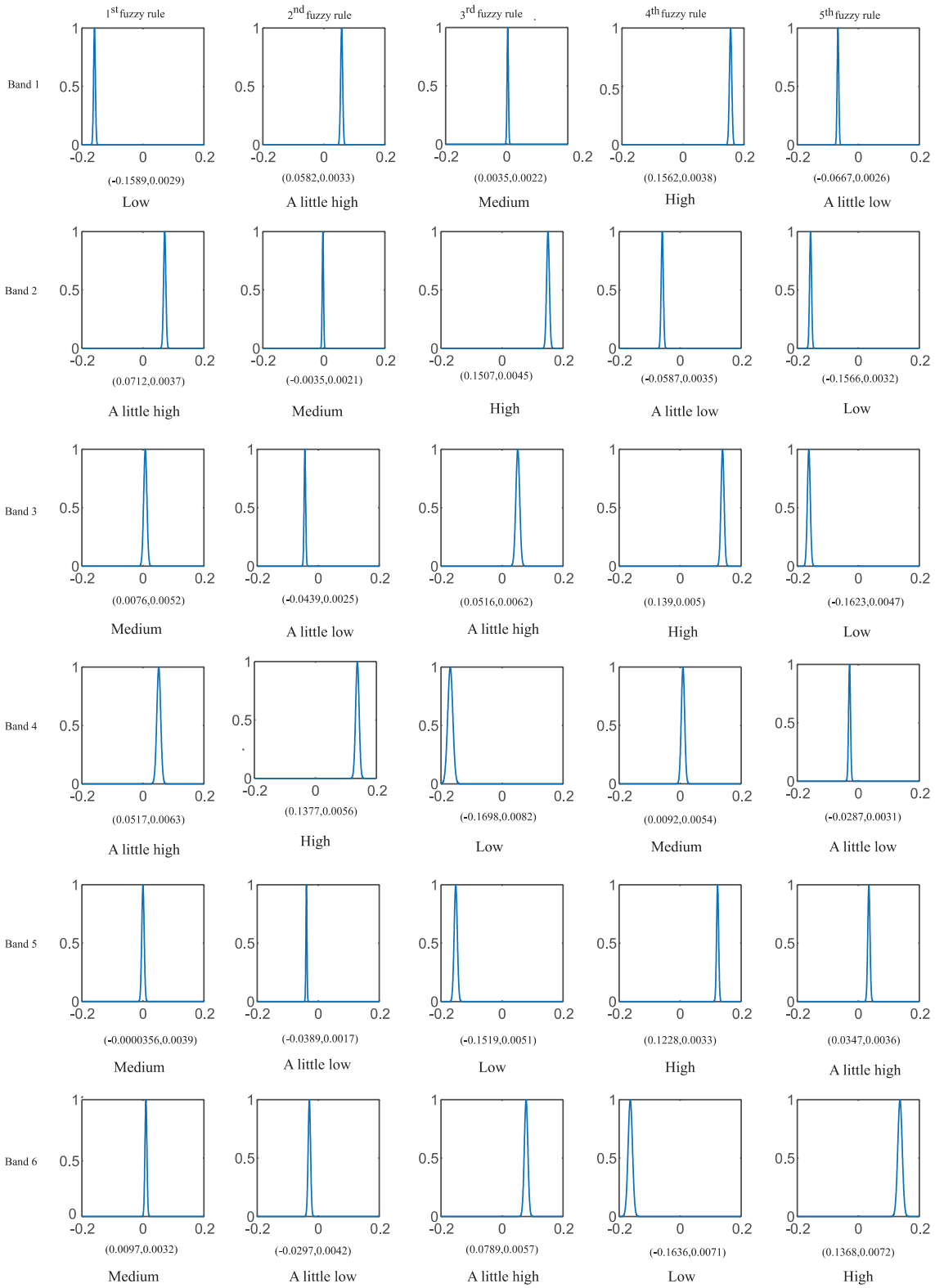


FIGURE 6 | The membership function of each fuzzy rule and the linguistic interpretation of each fuzzy subset of the TSK fuzzy system are obtained from the KPCA view.

of Holm's *post hoc* in **Table 8** also show that compared with other methods, TSK-TL—TL has better performance. This once again proves that our proposed TSK-TL has achieved better results in the detection of epileptic EEG signals.

Model Analysis

In this section, TSK-TL is analyzed through the model trained from the D2 dataset in the KPCA view. **Table 9** gives an example of the model, these model parameters prove that TSK-TL inherits the interpretability of the classical TSK fuzzy system. In this example, we set up five fuzzy rules. **Figure 6** shows the corresponding MF of each fuzzy set, where each MF has a fuzzy rule description, such as "the energy of a band of EEG signal is Low" (A little low, Medium, A little high or High). Because medical experts in different fields have different interpretations of fuzzy rules, the language description given is only a possible example.

According to the central value from low to high, these five MFs can be expressed as "Low," "A little low," "Medium," "A little high," and "High". Finally, through the language expression of the IF-part of the fuzzy rule and the linear function corresponding to the THEN-part of the fuzzy rule, five fuzzy rules of the KPCA view are given:

The first fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is Low, A little high, Medium, A little high, Medium, and Medium, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^1(\mathbf{x}) = [-2.7245-0.5118x_1 + 0.9243x_2-0.295x_3 \\ -0.0378x_4-0.58x_5-1.1201x_6]$$

The second fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is A little high, Medium, A little Low, High, A little Low and A little Low, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^2(\mathbf{x}) = [10.9842-8.5028x_1-0.9881x_2 + 4.7443x_3 \\ -1.8692x_4 + 3.2596x_5-3.0533x_6]$$

The third fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is Medium, High, A little high, Low, Low and A little high, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^3(\mathbf{x}) = [2.045 + 3.995x_1-4.3315x_2-1.4117x_3 \\ -0.9264x_4-0.0619x_5 + 0.4859x_6]$$

The fourth fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is High, A little low, High, Medium, High and Low, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^4(\mathbf{x}) = [-3.8877-2.2364x_1-0.7276x_2 \\ -1.1165x_3 + 0.1062x_4-1.3669x_5\ \ 0.9594x_6]$$

The fifth fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is A little Low, Low, Low, A little low, A little high, and High, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^5(\mathbf{x}) = [2.0489 + 4.0366x_1-4.2784x_2-1.4602x_3 \\ -1.0065x_4 + 0.0086x_5 + 0.9774x_6]$$

According to the final output value, i.e., y = -1 or y = 1, it can be judged whether the patient has epilepsy.

CONCLUSION

The study combines the classic TSK fuzzy system with transfer learning technology and proposes a TSK fuzzy system (TSK-TL) that is interpretable and can better adapt to scenarios of data distribution differences. It expands the application scenarios of the model and realizes the recognition of epileptic EEG signals with large data distribution differences in reality.

Although we have proved the effectiveness of TSK-TL, it can be further optimized. For example, there are several predefined parameters in the TSK-TL algorithm. When optimizing them, the optimization process of these parameters takes a lot of time. In the future, we will study the problem and develop more effective algorithms.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/benfulcher/hctsaTutorial_BonnEEG.

AUTHOR CONTRIBUTIONS

ZZ developed the theoretical framework and model in this work and drafted the manuscript. XD gave support for medical knowledge. ZZ, XD, JY, and AC implemented the algorithm and performed experiments and result analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grants U20A20228, 61772241, and 61873321, in part by the Science and Technology Demonstration Project of Social Development of Jiangsu Province under Grant BE2019631, and in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

REFERENCES

Aarabi, A., Fazel-Rezai, R., and Aghakhani, Y. (2009). A fuzzy rule-based system for epileptic seizure detection in intracranial EEG. *Clin. Neurophysiol.* 120, 1648–1657. doi: 10.1016/j.clinph.2009.07.002

Blanco, S., Kochen, S., Rosso, O. A., and Salgado, P. (1997). Applying time-frequency analysis to seizure EEG activity. *IEEE Engin. Med. Biol. Magaz.* 16, 64–71. doi: 10.1109/51.566156

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* 7, 1–30.

Deng, Z., Choi, K. S., Jiang, Y., and Wang, S. (2014b). Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods. *IEEE Trans. Cybernet.* 44, 2585–2599. doi: 10.1109/tcyb.2014.2311014

Deng, Z., Jiang, Y., Chung, F. L., Choi, K. S., and Wang, S. (2014a). Transfer prototype-based fuzzy clustering. *IEEE Trans. Fuzzy Syst.* 24, 1210–1232. doi: 10.1109/tfuzz.2015.2505330

Deng, Z., Jiang, Y., Chung, F. L., Ishibuchi, H., and Wang, S. (2012). Knowledge-leverage-based fuzzy system and its modeling. *IEEE Trans. Fuzzy Syst.* 21, 597–609. doi: 10.1109/tfuzz.2012.2212444

Deng, Z., Peng, X., Xie, L., Choi, K. S., and Wang, S. (2018). Transductive joint-knowledge-transfer tsk-fs for recognition of epileptic eeg signals. *IEEE Trans. Neural Syst. Rehabil. Engin.* 26, 1481–1494.

Dorai, A., and Ponnambalam, K. (2010). *Automated epileptic seizure onset detection. International Conference on Autonomous & Intelligent Systems*. New York, NY: IEEE.

Garcia, S., and Herrera, F. (2008). An extension on" Statistical Comparisons of Classifiers over Multiple Data Sets" for all pairwise comparisons. *J. Mach. Learn. Res.* 9, 2677–2694.

Iasemidis, L. D., Shiau, D. S., Chaovalitwongse, W., Sackellares, J. C., Pardalos, P. M., Principe, J. C., et al. (2003). Adaptive epileptic seizure prediction system. *IEEE Trans. Biomed. Engin.* 50, 616–627.

Iscan, Z., Dokur, Z., and Demiralp, T. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Syst. Appl.* 38, 10499–10505. doi: 10.1016/j.eswa.2011.02.110

Jiang, Y., Deng, Z., Chung, F. L., and Wang, S. (2015). Multi-task TSK fuzzy system modeling using inter-task correlation information. *Inform. Sci.* 298, 512–533. doi: 10.1016/j.ins.2014.12.007

Jiang, Y., Deng, Z., Chung, F. L., Wang, G., Qian, P., Choi, K. S., et al. (2017). Recognition of epileptic eeg signals using a novel multiview tsk fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/tfuzz.2016.2637405

Li, L. (2021). DSP for condition monitoring equipment in real-time. *ASP Trans. Neural Inform. Comput.* 1, 22–25.

Litt, B., Esteller, R., Echauz, J., D'Alessandro, M., Shor, R., Henry, T., et al. (2001). Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* 30, 51–64. doi: 10.1016/s0896-6273(01)00262-8

Long, M., Wang, J., Ding, G., Cheng, W., Zhang, X., and Wang, W. (2012). *Dual transfer learning. In Proceedings of the 2012 SIAM International Conference on Data Mining*. Thailand: Society for Industrial and Applied Mathematics, 540–551.

Long, M., Wang, J., Ding, G., Sun, J., and Yu, P. S. (2013). *Transfer feature learning with joint distribution adaptation. In Proceedings of the IEEE international conference on computer vision*. New York, NY: IEEE, 2200–2207.

Quanz, B., Huan, J., and Mishra, M. (2012). Knowledge transfer with low-quality data: A feature extraction issue. *IEEE Trans. Knowl. Data Engin.* 24, 1789–1802. doi: 10.1109/tkde.2012.75

Rabbi, A. F., and Fazel-Rezai, R. (2012). A fuzzy logic system for seizure onset detection in intracranial EEG. *Comput. Intell. Neurosci.* 12, 7051401–70514012.

Shi, X., Liu, Q., Fan, W., and Philip, S. Y. (2013). Transfer across completely different feature spaces via spectral embedding. *IEEE Trans. Knowl. Data Engin.* 25, 906–918. doi: 10.1109/tkde.2011.252

Srinivasan, V., Eswaran, C., and Sriraam, A. N. (2005). Artificial neural network based epileptic detection using time-domain and frequency-domain features. *J. Med. Syst.* 29, 647–660. doi: 10.1007/s10916-005-6133-1

Suk, K. H., Daesub, Y., Soon, S. H., and Hee, P. C. (2018). Predicting the eeg level of a driver based on driving information. *IEEE Trans. Intell. Transport. Syst.* 20, 1215–1225. doi: 10.1109/tits.2018.2848300

Tang, Y., and Durand, D. (2012). A tunable support vector machine assembly classifier for epileptic seizure detection. *Expert Syst. Appl.* 39, 3925–3938. doi: 10.1016/j.eswa.2011.08.088

Teng, T., Bi, L., and Liu, Y. (2017). Eeg-based detection of driver emergency braking intention for brain-controlled vehicles. *IEEE Trans. Intell. Transport. Syst.* 19, 1766–1773. doi: 10.1109/tits.2017.2740427

Tsujikawal, M., Onishil, Y., Kiuchil, Y., Ogatsul, T., and Hashimoto, S. (2018). *Drowsiness Estimation from Low-Frame-Rate Facial Videos using Eyelid Variability Features. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. New York, NY: IEEE.

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2009). Epileptic seizure detection in eegs using time–frequency analysis. *IEEE Trans. Inf. Technol. Biomed.* 13, 703–710. doi: 10.1109/titb.2009.2017939

Vivaldi, E. A., and Bassi, A. (2006). *Frequency domain analysis of sleep EEG for visualization and automated state detection. In 2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. New York: IEEE, 3740–3743.

Wang, C., and Mahadevan, S. (2011). *Heterogeneous domain adaptation using manifold alignment. In Twenty-second international joint conference on artificial intelligence*. Barcelona: IJCAI.

Wang, G., Deng, Z., and Choi, K. S. (2015). *Detection of Epileptic Seizures in EEG Signals with Rule-Based Interpretation by Random Forest Approach. International Conference on Intelligent Computing*. Cham: Springer.

Wang, J., Feng, W., Chen, Y., Yu, H., Huang, M., and Yu, P. S. (2018). *Visual domain adaptation with manifold embedded distribution alignment. In Proceedings of the 26th ACM international conference on Multimedia*. New York, NY: Association for Computing Machinery, 402–410.

Xiao, M., and Guo, Y. (2012). *Semi-supervised kernel matching for domain adaptation. In Twenty-Sixth AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press.

Yang, C., Deng, Z., Choi, K. S., Jiang, Y., and Wang, S. (2014). Transductive domain adaptive learning for epileptic electroencephalogram recognition. *Artific. Intell. Med.* 62, 165–177. doi: 10.1016/j.artmed.2014.10.002

Zhang, Z., Fan, C., Zhang, L., and Kong, M. (2021). K Algorithm for microstrip peocessor design. *ASP Trans. Neural Inform. Comput.* 1, 15–21.

Zhang, Z., Kawabata, H., and Liu, Z. Q. (2000). *EEG analysis using fast wavelet transform. In Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics.'cybernetics evolving to systems, humans, organizations, and their complex interactions'(cat. no. 0*, Vol. 4. New York: IEEE, 2959–2964.

Zheng, M. (2021). PCNN for power distribution network. *ASP Trans. Neural Inform. Comput.* 1, 9–14.

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fnins.2021.748689

Interpretable Recognition for Dementia Using Brain Images

*Xinjian Song[1], Feng Gu[2], Xiude Wang[2], Songhua Ma[3] and Li Wang[4,5,6]**

[1] Department of Rehabilitation Medicine, Affiliated Nantong Rehabilitation Hospital of Nantong University, Nantong, China, [2] Department of Medical Image, Affiliated Nantong Rehabilitation Hospital of Nantong University, Nantong, China, [3] Department of Neurology, Affiliated Nantong Rehabilitation Hospital of Nantong University, Nantong, China, [4] School of Information Science and Technology, Nantong University, Nantong, China, [5] Research Center for Intelligence Information Technology, Nantong University, Nantong, China, [6] Nantong Research Institute for Advanced Communication Technologies, Nantong, China

Machine learning-based models are widely used for neuroimage-based dementia recognition and achieve great success. However, most models omit the interpretability that is a very important factor regarding the confidence of a model. Takagi–Sugeno–Kang (TSK) fuzzy classifiers as the high interpretability and promising classification performance have widely used in many scenarios. TSK fuzzy classifier can generate interpretable fuzzy rules showing the reasoning process. However, when facing high-dimensional data, the antecedent become complex which may reduce the interpretability. In this study, to keep the antecedent of fuzzy rule concise, we introduce the subspace clustering technique and use it for antecedent learning. Experimental results show that the used model can generate promising recognition performance as well as concise fuzzy rules.

Keywords: dementia, Alzheimer's disease, brain images, TSK fuzzy systems, interpretability

OPEN ACCESS

Edited by:
Mohammad Khosravi, Persian Gulf University, Iran

Reviewed by:
Bing Li, Affiliated Cancer Hospital of Zhengzhou University, China
Yuexin Li, Hubei University, China

****Correspondence:***
Li Wang
wangli@ntu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *28 July 2021*
Accepted: *31 August 2021*
Published: *24 September 2021*

Citation:
Song X, Gu F, Wang X, Ma S and Wang L (2021) Interpretable Recognition for Dementia Using Brain Images. Front. Neurosci. 15:748689. doi: 10.3389/fnins.2021.748689

INTRODUCTION

Dementia is a clinical syndrome with progressive cognitive decline. The number of patients suffering from dementia worldwide is as high as 47.5 million. With the aging of the population, it is estimated that the number of people will be 75 million in another 20 years, and this number will triple in the next 50 years (Chen and Herskovits, 2010; Bansal et al., 2018). Alzheimer's Disease (AD) is the most common cause of dementia, which has a long incubation period and prodromal stage, and the average clinical treatment time is 8–10 years (Moradi et al., 2015; Zhang et al., 2015; Liu et al., 2020). There is currently no treatment that can stop, delay or reverse the progression of the course of AD. Neuropathological studies have found that the main causes of AD are the accumulation of amyloid plaques outside the cell, the tangling of neuronal fibers within the cell, the deterioration of synapses, and the death of neurons. The aggregation of amyloid plaques interferes with synaptic activity and brings about a series of inter-neural and intra-neuronal effects, and ultimately leads to the death of brain cells.

The current three-dimensional medical imaging technology is becoming more and more mature. Obtaining multiple modal medical images for each patient has become a diagnostic trend of AD. Such as complex but non-invasive magnetic resonance imaging (MRI) and positron emission tomography (PET) can realize the diagnosis of the disease and monitor its progress and the effect of subsequent treatment (Mirzaei et al., 2016; Zhang et al., 2021b). MRI is one of the neuroimaging modalities with high resolution imaging and high brain tissue contrast. It can well quantify the

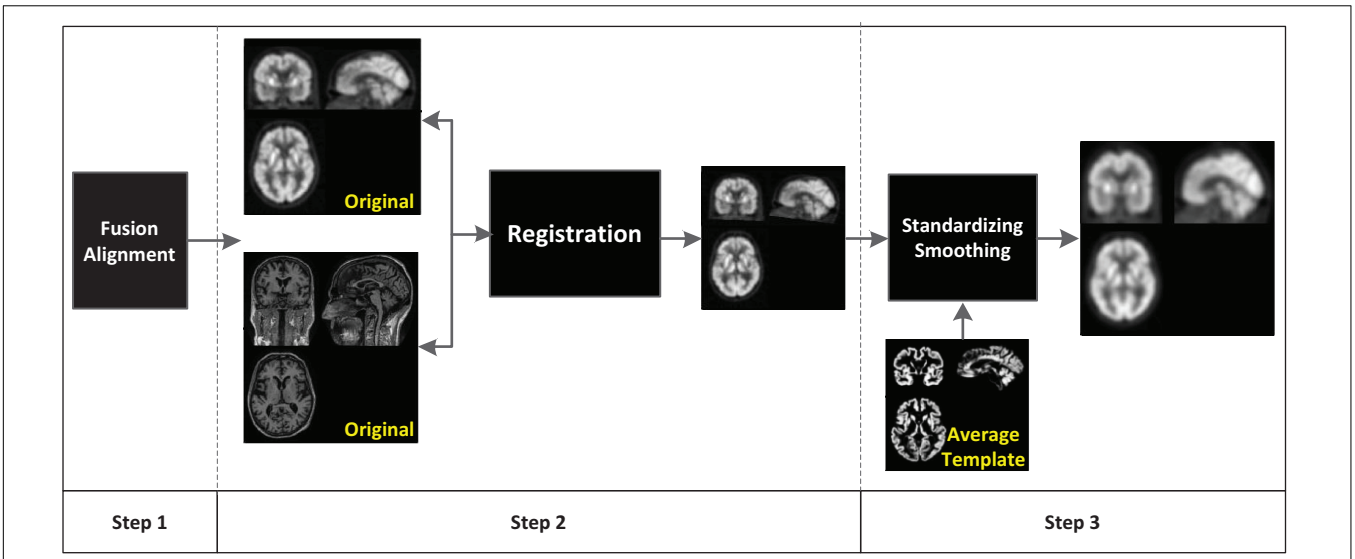


FIGURE 1 | Data preprocessing pipeline of positron emission tomography (PET) images.

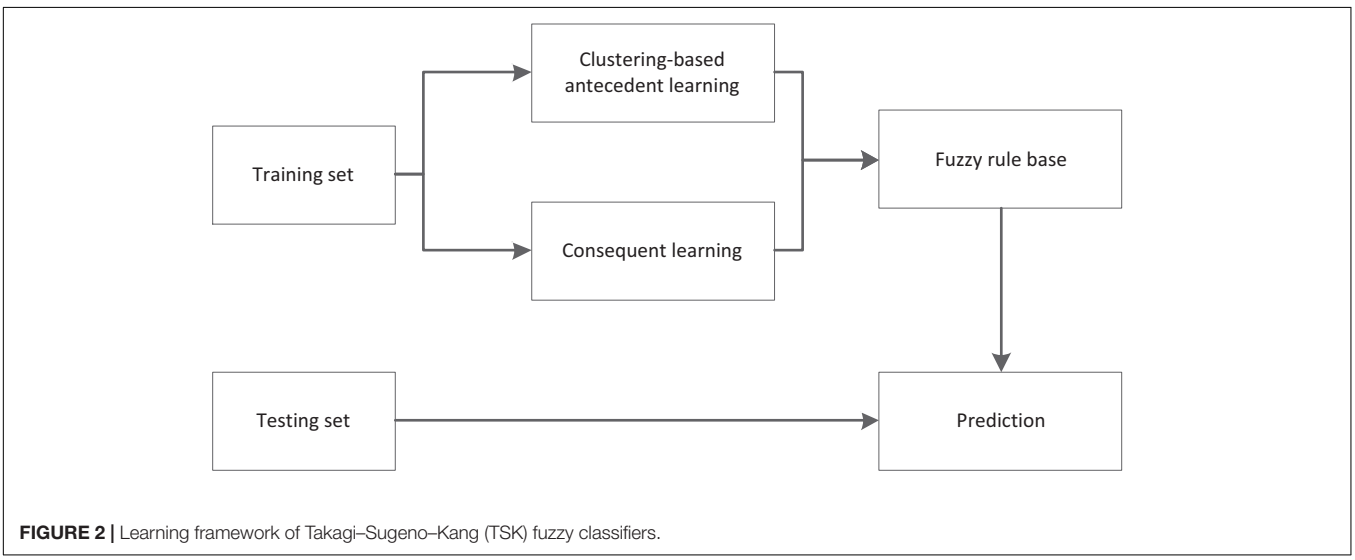


FIGURE 2 | Learning framework of Takagi–Sugeno–Kang (TSK) fuzzy classifiers.

brain tissue atrophy in patients with AD and mild cognitive impairment (MCI). PET is another neuroimaging modality for detecting AD. AD and MCI patients usually reduce glucose metabolism in certain areas before the brain is significantly atrophy. PET can monitor changes in glucose metabolism in the human body. In reality, the diagnosis of AD and MCI is still based on doctor's clinical diagnosis and psychometric evaluation. This method greatly wastes manpower and material resources, and at the same time produces highly subjective judgment results, which can easily lead to misdiagnosis and missed diagnosis. Patients with MCI will experience slight memory loss, but this will not have a substantial impact on the life of the patient. Therefore, the cognitive level of early MCI may not be judged according to the evaluation of the medical diagnosis cognitive scale. If you ignore it, then the risk of conversion to AD is extremely high, resulting in irreversible consequences, which is extremely detrimental to the early prevention of AD and MCI. Therefore, when looking for effective treatments to prevent or slow down the progress of AD, it is necessary to better develop medical auxiliary diagnostic tools, and the development of these tools also helps to measure the efficacy of new therapies.

Using machine learning methods to classify is to automatically learn the existing data, then obtain the corresponding patterns. Using such patterns, a set of unknown input samples can be judged to achieve classification and prediction. Machine learning methods have been widely used in character recognition, face recognition, speech recognition, and medical classification. Based on MRI, Cuingnet et al. (2011) compared 10 different AD automatic classification methods and compared the difference between extracting features of the whole brain and features of some related regions. The experiment proved that the effect of selecting a group of related regions is better than selecting the

whole brain. Area or separate hippocampus area. Querbes et al. (2009) used MRI images to measure the thickness of the cerebral cortex as a classification feature. The thickness of the cortex can characterize brain atrophy and achieved 85% classification accuracy in the classification of AD and HC. Wen et al. (2008) used principal component analysis to make feature selection for PET features, and then used logistic regression to classify AD and healthy controls (HC) and achieved a classification accuracy of 82%. Zhang et al. (2011) used support vector machine (SVM) to classify AD and HC based on multi-modal features and achieved a classification accuracy of 93.2%. Tong et al. (2017) used four modal features, namely MRI, PET, cerebrospinal fluid (CSF) and genetic information, and used an unsupervised metric fusion method based on cross-diffusion to perform feature fusion, and then classification of AD, MCI, and HC by Random Forest. The classification accuracy of AD and HC is 91.8%, and the classification of MCI and HC is 79.5%.

Although machine learning-based methods have been achieved great successful in recognition for dementia caused by AD, an important issue current models do not consider is the interpretability of a model. The interpretability of a model means that the model is not a black box, it has a mechanism to tell users how it works. Takagi–Sugeno–Kang (TSK) fuzzy classifiers as the high interpretability and promising classification performance have widely used in many scenarios (Visalakshi and Radha, 2014; Zhang et al., 2017; Jiang et al., 2020a; Xia et al., 2020). Compared with SVM (Zhang et al., 2021a), neural networks (NN), Random Forest, etc., TSK fuzzy classifiers are rule-based, and they can generate interpretable fuzzy rules which provide the evidence for the final classification results. However, TSK fuzzy classifiers are easy to suffer from "rule explosion" in the high-dimensional feature space. What's more, the high-dimensional feature space also leads to very complicated antecedents of fuzzy rules. Therefore, during the training phase, how to reduce irrelevant features is very important. To this end, in this study, we introduce a subspace clustering technique to the antecedent learning phase to ensure a concise antecedent of each fuzzy rule. The contributions of this study are summarized as follows.

(i) In order to keep the antecedents of fuzzy rules concise, a subspace clustering technique is introduced to reduce irrelevant features during antecedent learning.

(ii) We conduct extensive experiments to demonstrate the promising performance and good interpretability of our method.

DATA AND METHODS

Data

In this study, our brain PET images are provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) which is a 5-year public partnership sponsored by several institutes, companies, and non-profit organizations (Zhang et al., 2021b). **Figure 1** illustrates the data preprocessing pipeline of PET images, which can be divided into three main steps. In the first step, each subject in ADNI contains 96 PET images. Statistical parametric mapping (SPM) (Muzik et al., 2000) is used to fuse these PET images to construct a 3-D one which has brain spatial information and the feature information between tissue structures are also retained. In addition, motion correction is performed due to head motion. In the second step, the MRI image and PET image of each subject are registered, and affinely aligned. In the third step, the average template data generated in **Figure 1** is used to spatially normalize all PET images to the standard MNI space. PET images are also smoothed (8 mm Gaussian) to avoid the influences caused by noises.

The automated anatomical atlas (AAL; Rolls et al., 2020) which is available as a toolbox[1] for SPM is used as a template to extract original features from PET images. Based on AAL, the brain is segmented into 116 regions, and we select 90 regions from the cerebrum for feature extraction. To be specific, firstly, the PET images are resampled to the same size as the AAL template so that each region is in correspondence spatially. The size of AAL template is 61 × 73 × 61. Then we extract average intensity values from all regions of PET images as original features for our proposed classification model.

Methods

Figure 2 illustrates the learning framework of our TSK fuzzy classifier. The training contains two separate sections, clustering-based antecedent learning and consequent learning. In the following, we will focus on subspace clustering-based antecedent learning.

Notations

In this study, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in R^{N\times d}$ is used to represent the training sample set and $\mathrm{y} = [y_1, y_2, \ldots, y_n]^{\mathrm{T}} \in R^{n\times 1}$ is the corresponding label vector. An arbitrary sample $\mathbf{x}_i$ can be denoted as $[x_{i1}, x_{i2}, \ldots, x_{id}]^{\mathrm{T}}$. For an arbitrary matrix $\mathbf{B}$, we use b_{ij} to represent its element in the i-th row and j-th column and $\mathbf{b}_i$ to represent its i-th row.

Subspace Clustering-Based Takagi–Sugeno–Kang Fuzzy System

In this section, we develop a TSK fuzzy classifier to recognize AD patients. TSK fuzzy classifiers are rule-based models, the k-th fuzzy rule can be expressed as follows,

$$\textbf{If } x_{i1} \text{ is } A_1^k \ \wedge \ x_{i2} \textit{ is } A_2^k \ \wedge \ldots \wedge x_{id} \text{ is } A_d^k, \ \textbf{then } f^k(\mathbf{x}_i) = p_0^k + p_1^k x_{i1} + \cdots + p_d^k x_{id} \quad (1)$$

where A_i^k denotes the fuzzy subset regarding the i-th feature, $[p_0^k, p_1^k, \ldots, p_d^k]$ denotes the consequent parameter, $f^k(\mathbf{x}_i)$ denotes the output of the k-th fuzzy rule regarding $\mathbf{x}_i$. When we adopt multiplication as conjunction and implication, addition as combination, and the center of gravity as defuzzification, the output of the TSK fuzzy classifier can be expressed as follows,

$$y(\mathbf{x}_i) = \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x}_i)}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x}_i)} f^k(\mathbf{x}_i) = \sum_{k=1}^{K} \tilde{\mu}^k(\mathbf{x}_i) f^k(\mathbf{x}_i) \quad (2)$$

[1] http://www.gin.cnrs.fr/AAL

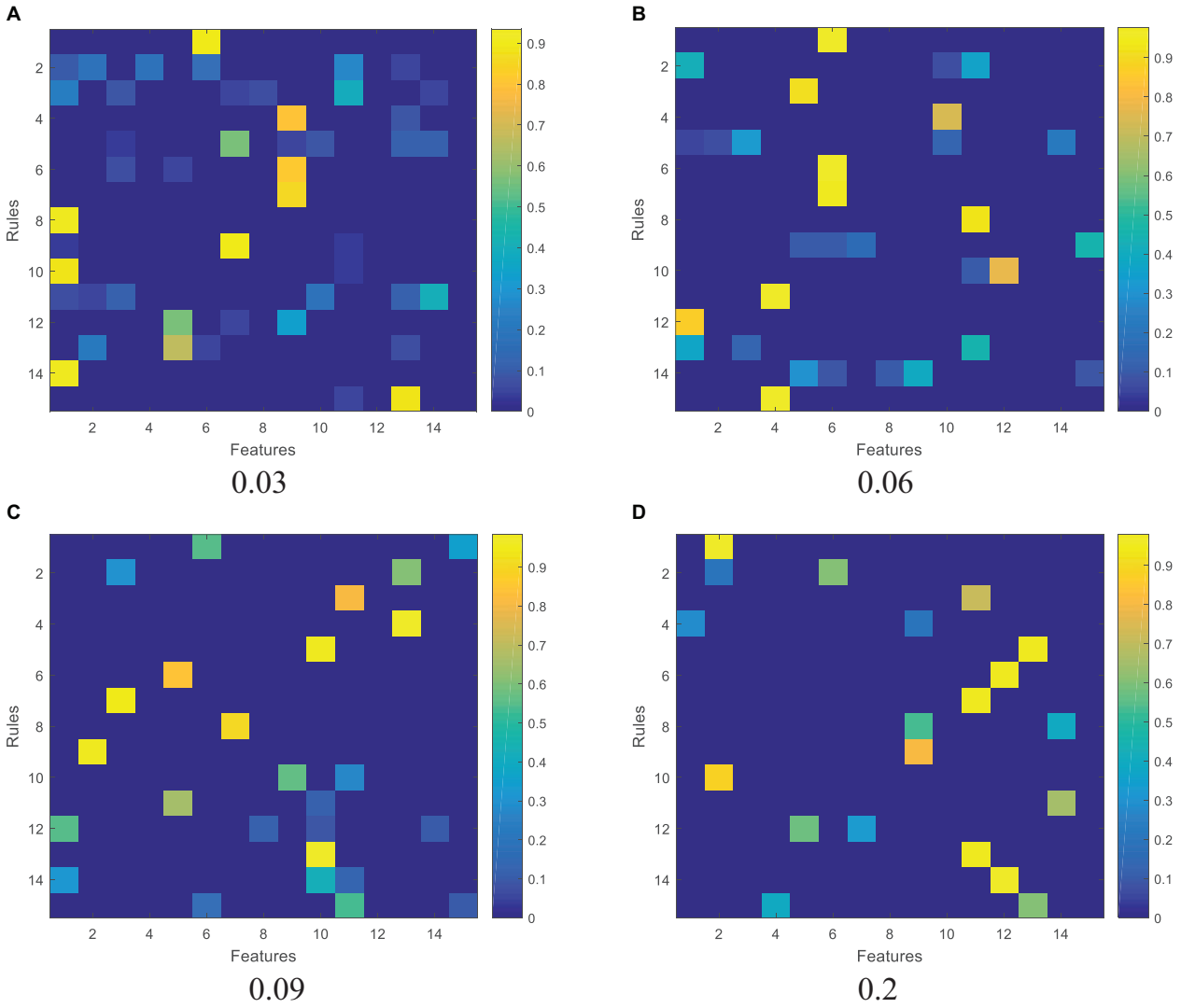


FIGURE 3 | Activation of features by the subspace clustering under different thresholds: **(A)** 0.03, **(B)** 0.06, **(C)** 0.09, and **(D)** 0.2.

where K denotes the number of fuzzy rules, $\mu^k(\mathbf{x}_i)$ and $\tilde{\mu}^k(\mathbf{x}_i)$ are usually called as the firing strength and the normalized firing strength, respectively, which are defined as follows,

$$\mu^k(\mathbf{x}_\mathrm{i}) = \prod_{j=1}^{d} \mu_{A_i^k}(\mathbf{x}_\mathrm{i}) \tag{3}$$

$$\tilde{\mu}^k(\mathbf{x}_\mathrm{i}) = \frac{\mu^k(\mathbf{x}_\mathrm{i})}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x}_\mathrm{i})} \tag{4}$$

where $\mu_{A_i^k}(\mathbf{x}_i)$ denotes the membership function the fuzzy subset A_i^k. In this study, we adopt the Gaussian function as the membership function, which is defined as follows,

$$\mu_{A_i^k}(\mathbf{x}_\mathrm{i}) = \exp\left(\frac{-(\mathbf{x}_\mathrm{i} - \mathbf{v}_\mathrm{i}^\mathrm{k})}{2\sigma_i^k}\right) \tag{5}$$

where $\mathbf{v}_i^k$ and σ_i^k are the antecedent parameters.

Once the antecedent parameters are determined clustering techniques or other schemas, let

$$\mathbf{x}_\mathrm{e} = \left(1, \mathbf{x}_i^T\right)^T \tag{6}$$

$$\tilde{\mathbf{x}}^k = \tilde{\mu}^k(\mathbf{x}_\mathrm{i})\,\mathbf{x}_\mathrm{e} \tag{7}$$

$$\mathbf{x}_g = \left(\left(\tilde{\mathbf{x}}^1\right)^T, \left(\tilde{\mathbf{x}}^2\right)^T, \ldots, \left(\tilde{\mathbf{x}}^K\right)^T\right)^T \tag{8}$$

$$\mathbf{p}^k = \left(\mathbf{p}_0^k, \mathbf{p}_1^k, \ldots, \mathbf{p}_d^k\right)^T \tag{9}$$

$$\mathbf{p}_g = \left(\left(\mathbf{p}^1\right)^T, \left(\mathbf{p}^2\right)^T, \ldots, \left(\mathbf{p}^K\right)^T\right)^T \tag{10}$$

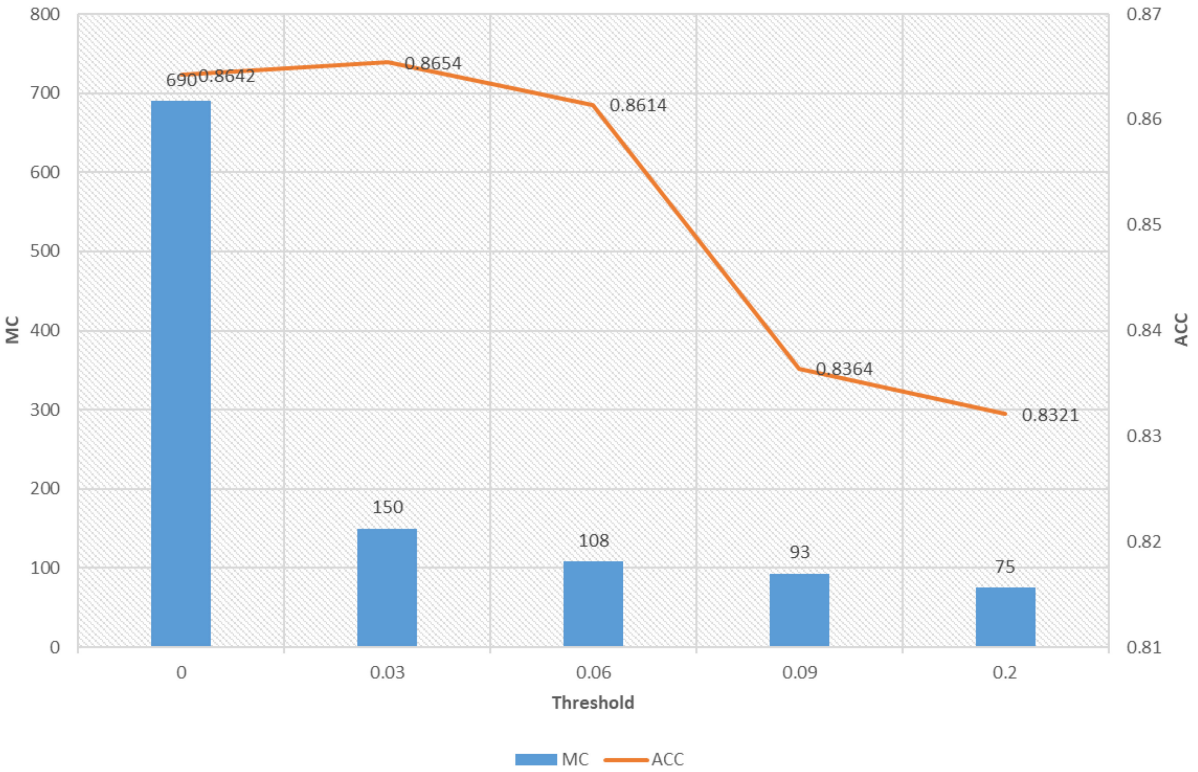


FIGURE 4 | Model complexity and accuracy under different thresholds.

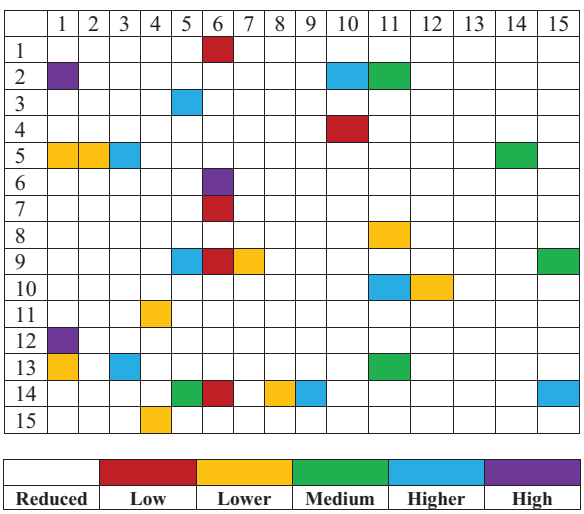


FIGURE 5 | Linguistic meaning of activated features of each rule.

Based on (6)–(10), we can update the output of the TSK fuzzy classifier as follows,

$$y(\mathbf{x}_i) = \mathbf{p}_g^T \mathbf{x}_g \tag{11}$$

In general, the optimization of the TSK fuzzy classifier can be conduct separately. As for the antecedent, clustering is usually used, and for the consequent, we see from (11) that it can be solved by many techniques because it can be considered as a linear regression model. As we stated before that the number of features involved in antecedents of fuzzy rules is a key factor to the interpretability of TSK fuzzy systems. Therefore, to reduce irrelevant features and make the antecedents of fuzzy rules more concise, in our study, we introduce a subspace clustering technique to optimize the antecedent. The core idea is that it uses a weight matrix to measure the weights of features in each cluster. The objective function of the introduced clustering technique is formulated as follows,

$$J(\mathbf{U}, \mathbf{V}, \mathbf{W}) = \sum_{c=1}^{C} \sum_{i=1}^{N} \mu_{ci}^{m} \sum_{j=1}^{d} w_{cj} \left(x_{ij} - v_{cj}\right)^2 + \sum_{c=1}^{c} \delta_c \sum_{j=1}^{d} w_{cj}^2, \tag{12}$$

$$\text{s.t.} \sum_{c=1}^{C} \mu_{ci} = 1, \quad \sum_{j=1}^{d} w_{cj} = 1 \tag{13}$$

where μ_{ci} is an element of $\mathbf{U}$ which denotes the fuzzy membership degree of sample $\mathbf{x}_i$ belonging to cluster c, v_{cj} is an element of $\mathbf{V}$ which denotes the j-th feature of the c-th cluster's center, and w_{cj} is an element of $\mathbf{W}$ which denotes the weight of the j-th feature in the c-th cluster. δ_c is constant of the c-th cluster, C denotes the number of clusters, N denotes the number of training samples, d denotes the number of features and m denotes the fuzzy exponential.

TABLE 1 | Rule base.

Fuzzy rule: If x_{i1} is $A_1^k \wedge x_{i2}$ is $A_2^k \wedge \ldots \wedge x_{id}$ is A_d^k, **then** $f^k(\mathbf{x_i}) = p_0^k + p_1^k x_{i1} + \cdots + p_d^k x_{id}$		
Rule No.	**Antecedent**	**Consequent** $\left[p_0^k, p_1^k, \ldots, p_d^k\right]$
1	**If** x_{i6} is *Low*	[0.0558, −0.1106, −0.1158, 0.0490, −0.0513, 0.0706, 0.0274, −0.0270, 0.1006, 0.0865, 0.0762, 0.0642, 0.0642, −0.0763, 0.2249, −0.1551]
2	**If** x_{i1} is *High* ∧ x_{i10} is *Higher* ∧ x_{i11} is *Medium*	[−0.1494, 0.0843, −0.1473, −0.2774, 0.1236, −0.1115, 0.0795, 0.0231, −0.0210, 0.0027, 0.0950, 0.0711, 0.0687, 0.0585, −0.0767, 0.2036]
3	**If** x_{i1} is *High* ∧ x_{i10} is *Higher* ∧ ∧x_{i11} is *Medium*	[0.1820, −0.1378, −0.2689, −0.2122, 0.1615, −0.0821, −0.1418, 0.0769, 0.0151, 0.2019, 0.1007, 0.0927, 0.0750, 0.0784, 0.0609, −0.0758]
4	**If** x_{i5} is *Higher*	[−0.0708, 0.1903, −0.0854, −0.1610, 0.0968, −0.0573, −0.1169, −0.1359, 0.0722, −0.0805, −0.0210, 0.0912, 0.0944, 0.0584, 0.0850, 0.0562]
5	**If** x_{i10} is *Lower*	[0.0533, −0.0488, −0.0385, −0.1871, 0.0617, −0.1196, −0.1453, −0.1122, −0.1232, 0.0722, 0.0151, −0.0153, 0.0985, 0.1007, 0.0562, 0.0723]
6	**If** x_{i6} is *High*	[0.0607, 0.0716, 0.2803, 0.4842, −0.1474, 0.1020, −0.1651, −0.1353, −0.1002, −0.0581, 0.0722, 0.0122, −0.0199, 0.0505, 0.1021, 0.0601]
7	**If** x_{i6} is *Low*	[0.0655, 0.0750, 0.0382, −0.0501, 0.1809, −0.1283, 0.0952, −0.1584, −0.1041, 0.1218, −0.1232, 0.0708, 0.0144, 0.0629, 0.0448, 0.0987]
8	**If** x_{i11} is *Lower*	[0.0922, 0.0798, 0.0277, 0.0025, −0.0677, 0.2082, −0.1321, 0.0787, −0.1413, 0.2391, −0.1002, −0.1169, 0.0718, 0.0002, 0.0808, 0.0564]
9	**If** x_{i5} is *Higher* ∧ x_{i6} is *Low* ∧ x_{i7} is *Lower* ∧ x_{i15} is *Medium*	[0.0776, 0.0968, 0.1594, 0.0372, 0.0549, −0.0278, 0.2004, −0.1404, −0.0509, 0.0285, −0.1043, −0.0917, −0.1218, 0.0821, −0.0053, 0.0459]
10	**If** x_{i11} is *Higher* ∧ x_{i12} is *Lower*	[−0.0004, 0.1066, 0.5270, −0.2723, 0.0613, 0.0965, −0.0364, 0.1864, −0.1640, 0.0788, −0.1414, −0.0815, −0.0985, −0.0874, 0.0844, 0.0041]
11	**If** x_{i4} is *Lower*	[0.0096, −0.0235, 0.1841, −0.0057, 0.0676, 0.0958, 0.0857, −0.0543, 0.3149, −0.1403, 0.0697, −0.1306, −0.0995, −0.0062, −0.0831, 0.0793]
12	**If** x_{i1} is *High*	[0.0713, 0.0174, 0.0144, −0.2364, 0.0919, 0.0945, 0.0868, 0.0660, −0.0819, 0.1864, −0.1443, 0.0767, −0.1391, 0.0844, 0.0113, −0.0919]
13	**If** x_{i1} is *Lower* ∧ x_{i3} is *Higher* ∧ x_{i1} is *Medium*	[−0.1075, 0.0735, −0.2035, 0.2646, 0.0833, 0.1043, 0.0883, 0.0702, 0.0704, −0.0543, 0.1821, −0.1413, 0.0245, −0.0569, 0.1103, −0.0234]
14	**If** x_{i5} is *Medium* ∧ x_{i6} is *Low* ∧ x_{i8} is *Lower* ∧ x_{i9} is *Higher* ∧ x_{i15} is *Higher*	[−0.0733, −0.1272, −0.2476, −0.3230, −0.0079, 0.1286, 0.1012, 0.0762, 0.1231, 0.0661, −0.0622, 0.1852, −0.1563, 0.0165, −0.0448, 0.0573]
15	**If** x_{i5} is *Lower*	[−0.0368, −0.1045, 0.0095, −0.2519, 0.0104, −0.0287, 0.1196, 0.0949, 0.0234, 0.0702, 0.0589, −0.0562, 0.2137, −0.1574, 0.0320, −0.0693]

According to Frigui and Nasraoui (2004), by introducing Lagrangian multipliers, we have several updating rules as follows,

$$w_{cj} = \frac{1}{d} + \frac{1}{2\delta_c}\sum_{i=1}^{N}\mu_{ci}^m\left(\frac{||x_i - v_c||^2}{d} - \left(x_{ij} - v_{cj}\right)^2\right), \quad (14)$$

$$\delta_c = \frac{\sum_{i=1}^{N}\mu_{ci}^m\sum_{j=1}^{d}w_{cj}(\mathrm{x}_{ij} - \mathrm{v}_{cj})^2}{\sum_{j=1}^{d}w_{cj}^2} \quad (15)$$

$$\mu_{ci} = \frac{1}{\sum_{c'=1}^{C}\left[\frac{\sum_{j=1}^{d}w_{cj}(x_{ij}-v_{cj})^2}{\sum_{j=1}^{d}w_{c'j}(x_{ij}-v_{c'j})^2}\right]^{1/(m-1)}} \quad (16)$$

$$v_{cj} = \frac{w_{cj}\sum_{i=1}^{N}\mu_{ci}^m x_{ij}}{w_{cj}\sum_{i=1}^{N}\mu_{ci}^m} \quad (17)$$

When the subspace clustering converges, we can use the following equations to calculate the antecedent parameters $\mathbf{V}_i^k$ and σ_i^k,

$$\mathbf{v}_i^k = \frac{\sum_{i=1}^{N}\mu_{ki}\mathbf{x}_{ij}}{\sum_{i=1}^{N}\mu_{ki}}, \quad (18)$$

$$\sigma_i^k = \frac{h\sum_{i=1}^{N}\mu_{ki}(\mathbf{x}_{ij} - \mathbf{v}_i^k)^2}{\sum_{i=1}^{N}\mu_{ki}}, \quad (19)$$

where h is a user-defined parameter. Based on the subspace clustering technique, the training algorithm of the TSK fuzzy classifier is listed as follows. Notably, the stopping threshold ε is set to 1e-5. Detailed algorithm steps are shown in **Algorithm 1**.

Algorithm 1 | Subspace clustering-based TSK fuzzy system.

Input:
$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, $y = [y_1, y_2, \ldots, y_n]^T$, number of fuzzy rules K, fuzzy exponential m
Output:
Antecedent parameters v_i^k and σ_i^k
Consequent parameters $\mathbf{p}_g^T$
Procedure:
Stage-1: Antecedent learning
[1]Initialize $\mathbf{U}(0)$
[2]Initialize V(0)
[3]Initialize W(0)
[4]Set $t \leftarrow 1$
Repeat
[5]Update $\mathbf{U}(t)$ by using (16)
[6]Update $\mathbf{V}(t)$ by using (17)
[7]Update $\mathbf{W}(t)$ by using (14)
[8]Update $\delta_c(t)$ by using (15)
[9] $t \leftarrow t + 1$
Until $||\mathbf{U}(t+1) - \mathbf{U}(t)||^2 < \varepsilon$
[10]Use (18) and (19) to compute v_i^k and σ_i^k
Stage-2: Consequent learning
[1]Use (6)–(10) to get $\mathbf{p}_g^T$ and $\mathbf{x}_g$
[2]Construct a quadratic programming problem and solve it to get $\mathbf{p}_g^T$

RESULTS

Setups

In our experiments, the fuzzy exponential m is set to 2, the number of fuzzy rules is set to 15, h in (19) is set to 0.5. The original number of features we obtained via the pipeline in **Figure 1** is 93. We use the feature selection method proposed in Jiang et al. (2020b) to reduce the dimension to 15.

To highlight the interpretability and performance of the subspace-based TSK fuzzy classifier, we introduce the classical one order TSK fuzzy classifier (1-TSK-FC) (Jiang et al., 2016) for comparison.

We introduce accuracy (ACC) and model complexity (MC) to evaluate the performance and interpretability, where ACC is defined as the ratio of correctly classified samples to the total number of samples, and MC is defined as the number of parameters participating the training phase.

Experimental Results

We report the experimental results from 3 aspects. The first one is the feature activation results, as shown in **Figure 3**, regarding the subspace clustering for antecedent learning. In **Figure 3**, each subpanel represents the activated features for each fuzzy rule under different thresholds, the brighter the color, the greater the corresponding weight of each feature in each fuzzy rule. It observes that as the threshold increases, the number of activated features contained in each rule begins to decrease.

The second one is the relationship between model complexity and accuracy, which is illustrated in **Figure 4**. As we stated before, model complexity can be quantificationally measured by the involved number of parameters during antecedent learning and consequent learning. For example, when the threshold is set to 0.06, based on the feature reduction result shown in **Figure 3B**, the number of features involved in each feature is 1, 3, 1, 1, 5, 1, 1, 1, 4, 2, 1, 1, 3, 5, and 1, respectively. According to (5), we know that each feature needs two parameters, so, during the antecedent learning phase, the number of parameters each feature needs is 2, 6, 2, 2, 10, 2, 2, 2, 8, 4, 2, 2, 6, 10, and 2, respectively. During the phase of consequent learning, according to (1), we know that each feature needs $d + 1$ parameters, where d is the current dimension after feature reduction. That is, each feature needs 2, 4, 2, 2, 6, 2, 2, 2, 5, 3, 2, 2, 4, 6, and 2 parameters, respectively. Therefore, model complexity under threshold being 0.06 is 108. When the threshold is set to 0, it means that the classifier degenerates into 1-TSK-FS. From **Figure 4**, it observes that model complexity of 1-TSK-FS is 690, which is seriously higher than that of subspace clustering-based learning. What is more, the classification performance does not reduce significantly with the decreasing of model complexity. For example, when the model complexity is 75, the corresponding performance still keeps in a reasonable level.

The third one is the results of model interpretability. In **Figure 5**, we assign linguistic terms "*Low*, *Lower*, *Medium*, *Higher*, and *High*" to each feature according to the antecedent parameters. Based on this assignment and the consequent parameters, **Table 1** shows the rule base consisting of 15 fuzzy rules. It is easy to find that the antecedent of each fuzzy rule is very concise. Please note that the assignment of linguistic terms is based on the knowledge of expert. Different experts from different domain may have different assignment.

DISCUSSION

Although there have many excellent models that can be used for AD detection based on neuroimages, most of them omit

the interpretability that is a very important factor regarding the confidence of a model. TSK fuzzy systems are rule-based inference models which can illustrate the reasoning process of the generated results. Therefore, owning to the high interpretability, they are widely used in many application scenarios. In this study, we introduce a subspace clustering technique and embed it into the antecedent learning phase to address the issue of rule complexity caused by the high-dimensional input feature space.

The subspace clustering technique uses a weighting strategy to measure the weight of each feature in each cluster. We know that when the clustering technique is used for antecedent learning of TSK fuzzy systems, the number of clusters is set to the number of fuzzy rules. Hence, the weight of each feature in each cluster corresponds to the compatible degree of each feature in each fuzzy rule. In this study, we define a threshold to reduce the irrelevant feature to keep the antecedent concise.

Definitely, we can use different thresholds to control the feature distribution. From **Figure 3**, we can find that the greater the threshold, the sparser distribution of the features in each fuzzy rule. In theory, the fewer features, the more succinct the antecedent of the rule, and therefore the stronger the interpretability of the fuzzy rule. However, too few features will affect the reasoning process and thus affect the classification accuracy. As can be seen from **Figure 3** that when the threshold is set from 0.06 to 0.2, the classification performance in terms of accuracy decreases from 0.8614 to 0.8321. Therefore, the threshold should be elastically set to keep the balance between classification performance and interpretability.

Overall, from the experimental results, we find that subspace clustering-based TSK fuzzy classifiers cannot only ensure promising performance but also guarantee concise antecedents of fuzzy rules. Compared with classical clustering methods, like fuzzy c-means (FCM), our method is more flexible.

CONCLUSION

In this study, we employ an interpretable model to achieve the detection of AD patients based on neuroimages. Compared with existing models, it merits lie in that it can generate fuzzy rules for reasoning. What's more, we introduce a subspace clustering technique to keep the fuzzy rule concise. In our future work, we can design more strategies to reduce the superfluous fuzzy rules to further improve the interpretability of the model.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://adni.loni.usc.edu/about/.

AUTHOR CONTRIBUTIONS

XS, FG, XW, and SM contributed on data preprocessing. LW contributed on coding and writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partly supported by National Science Foundation of China (No. 81873915).

ACKNOWLEDGMENTS

We thank the reviewers whose comments and suggestions helped improve this manuscript.

REFERENCES

Bansal, D., Chhikara, R., Khanna, K., and Gupta, P. (2018). Comparative analysis of various machine learning algorithms for detecting dementia. *Procedia Comput. Sci.* 132, 1497–1502. doi: 10.1016/j.procs.2018.05.102

Chen, R., and Herskovits, E. H. (2010). Machine-learning techniques for building a diagnostic model for very mild dementia. *Neuroimage* 52, 234–244. doi: 10.1016/j.neuroimage.2010.03.084

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M. O., et al. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 56, 766–781. doi: 10.1016/j.neuroimage.2010.06.013

Frigui, H., and Nasraoui, O. (2004). Unsupervised learning of prototypes and attribute weights. *Pattern Recognit.* 37, 567–581. doi: 10.1016/j.patcog.2003.08.002

Jiang, K., Tang, J., Wang, Y., Qiu, C., Zhang, Y., and Lin, C. (2020b). EEG feature selection via stacked deep embedded regression with joint sparsity. *Front. Neurosci.* 14:829. doi: 10.3389/fnins.2020.00829

Jiang, Y., Deng, Z., Chung, F. L., Wang, G., Qian, P., Choi, K. S., et al. (2016). Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/tfuzz.2016.2637405

Jiang, Y., Zhang, Y., Lin, C., Wu, D., and Lin, C. T. (2020a). EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system. *IEEE Trans. Intell. Transp. Syst.* 22, 1752–1764. doi: 10.1109/tits.2020.2973673

Liu, L., Zhao, S., Chen, H., and Wang, A. (2020). A new machine learning method for identifying Alzheimer's disease. *Simul. Model. Pract. Theory* 99:102023. doi: 10.1016/j.simpat.2019.102023

Mirzaei, G., Adeli, A., and Adeli, H. (2016). Imaging and machine learning techniques for diagnosis of Alzheimer's disease. *Rev. Neurosci.* 27, 857–870. doi: 10.1515/revneuro-2016-0029

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., and Alzheimer's Disease Neuroimaging Initiative (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 104, 398–412. doi: 10.1016/j.neuroimage.2014.10.002

Muzik, O., Chugani, D. C., Juhász, C., Shen, C., and Chugani, H. T. (2000). Statistical parametric mapping: assessment of application in children. *Neuroimage* 12, 538–549. doi: 10.1006/nimg.2000.0651

Querbes, O., Aubry, F., Pariente, J., Lotterie, J. A., Démonet, J. F., Duret, V., et al. (2009). Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* 132, 2036–2047. doi: 10.1093/brain/awp105

Rolls, E. T., Huang, C. C., Lin, C. P., Feng, J., and Joliot, M. (2020). Automated anatomical labelling atlas 3. *Neuroimage* 206:116189. doi: 10.1016/j.neuroimage.2019.116189

Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D., and Alzheimer's Disease Neuroimaging Initiative (2017). Multi-modal classification of Alzheimer's

disease using nonlinear graph fusion. *Pattern Recognit.* 63, 171–181. doi: 10.1016/j.patcog.2016.10.009

Visalakshi, S., and Radha, V. (2014). "A literature review of feature selection techniques and applications: review of feature selection in data mining," in *Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, 1–6.

Wen, L., Bewley, M., Eberl, S., Fulham, M., and Feng, D. (2008). "Classification of dementia from FDG-PET parametric images using data mining," in *Proceedings of the 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, (Piscataway, NJ: IEEE), 412–415.

Xia, K., Zhang, Y., Jiang, Y., Qian, P., Dong, J., Yin, H., et al. (2020). TSK fuzzy system for multi-view data discovery underlying label relaxation and cross-rule & cross-view sparsity regularizations. *IEEE Trans. Industr. Inform.* 17, 3282–3291. doi: 10.1109/tii.2020.3007174

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867.

Zhang, Y., Dong, Z., Phillips, P., Wang, S., Ji, G., Yang, J., et al. (2015). Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Front. Comput. Neurosci.* 9:66. doi: 10.3389/fncom.2015.00066

Zhang, Y., Ishibuchi, H., and Wang, S. (2017). Deep Takagi–Sugeno–Kang fuzzy classifier with shared linguistic fuzzy rules. *IEEE Trans. Fuzzy Syst.* 26, 1535–1549. doi: 10.1109/tfuzz.2017.2729507

Zhang, Y., Wang, G., Chung, F. L., and Wang, S. (2021a). Support vector machines with the known feature-evolution priors. *Knowl. Based Syst.* 223:107048. doi: 10.1016/j.knosys.2021.107048

Zhang, Y., Wang, S., Xia, K., Jiang, Y., Qian, P., and Alzheimer's Disease Neuroimaging Initiative (2021b). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inf. Fusion* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fnins.2021.744296

Improvement Using Planomics Features on Prediction and Classification of Patient-Specific Quality Assurance Using Head and Neck Volumetric Modulated Arc Therapy Plan

*Bing Li[1†], Junying Chen[2†], Wei Guo[1], Ronghu Mao[1], Xiaoli Zheng[1], Xiuyan Cheng[1], Tiantian Cui[1], Zhaoyang Lou[1], Ting Wang[1], Dingjie Li[1], Hongyan Tao[3], Hongchang Lei[1] and Hong Ge[1]**

[1] Department of Radiation Oncology, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China, [2] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China, [3] Department of Planning and Finance, Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Wenzheng Sun,
Zhejiang University, China
Bin Liang,
Center for Cancer Research, National Cancer Institute, United States

***Correspondence:**
Hong Ge
gehong616@126.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 20 July 2021
Accepted: 17 August 2021
Published: 01 October 2021

Citation:
Li B, Chen J, Guo W, Mao R, Zheng X, Cheng X, Cui T, Lou Z, Wang T, Li D, Tao H, Lei H and Ge H (2021) Improvement Using Planomics Features on Prediction and Classification of Patient-Specific Quality Assurance Using Head and Neck Volumetric Modulated Arc Therapy Plan. Front. Neurosci. 15:744296. doi: 10.3389/fnins.2021.744296

Purpose: This study aimed to evaluate the utility of a new plan feature (planomics feature) for predicting the results of patient-specific quality assurance using the head and neck (H&N) volumetric modulated arc therapy (VMAT) plan.

Methods: One hundred and thirty-one H&N VMAT plans in our institution from 2019 to 2021 were retrospectively collected. Dosimetric verification for all plans was carried out using the portal dosimetry system integrated into the Eclipse treatment planning system based on the electronic portal imaging devices. Gamma passing rates (GPR) were analyzed using three gamma indices of 3%/3 mm, 3%/2 mm, and 2%/2 mm with a 10% dose threshold. Forty-eight conventional features affecting the dose delivery accuracy were used in the study, and 2,476 planomics features were extracted based on the radiotherapy plan file. Three prediction and classification models using conventional features (CF), planomics features (PF), and hybrid features (HF) combining two sets of features were constructed by the gradient boosting regressor (GBR) and Ridge classifier for each GPR of 3%/3 mm, 3%/2 mm, and 2%/2 mm, respectively. The absolute prediction error (APE) and the area under the curve (AUC) were adopted for assessing the performance of prediction and classification models.

Results: In the GPR prediction, the average APE of the models using CF, PF, and HF was 1.3 ± 1.2%/3.6 ± 3.0%, 1.7 ± 1.5%/3.8 ± 3.5%, and 1.1 ± 1.0%/4.1 ± 3.1% for 2%/2 mm; 0.7 ± 0.6%/2.0 ± 2.0%, 1.0 ± 1.1%/2.2 ± 1.8%, and 0.6 ± 0.6%/2.2 ± 1.9% for 3%/2 mm; and 0.4 ± 0.3%/1.2 ± 1.2%, 0.4 ± 0.5%/1.3 ± 1.0%, and 0.3 ± 0.3%/1.2 ± 1.1% for 3%/3 mm, respectively. In the regression prediction, three models give a similar modeling performance for predicting the GPR. The classification results were 0.67 ± 0.03/0.66 ± 0.07, 0.77 ± 0.03/0.73 ± 0.06, and

0.78 ± 0.02/0.75 ± 0.04 for 3%/3 mm, respectively. For 3%/2 mm, the AUCs of the training and testing cohorts were 0.64 ± 0.03/0.62 ± 0.07, 0.70 ± 0.03/0.67 ± 0.06, and 0.75 ± 0.03/0.71 ± 0.07, respectively, and for 2%/2 mm, the average AUCs of the training and testing cohorts were 0.72 ± 0.03/0.72 ± 0.06, 0.78 ± 0.04/0.73 ± 0.07, and 0.81 ± 0.03/0.75 ± 0.06, respectively. In the classification, the PF model has a better classification performance than the CF model. Moreover, the HF model provides the best result among the three classifications models.

Conclusions: The planomics features can be used for predicting and classifying the GPR results and for improving the model performance after combining the conventional features for the GPR classification.

Keywords: VMAT, H&N, quality assurance, radiotherapy, machine learning

INTRODUCTION

Volumetric modulated arc therapy (VMAT) is advanced radiotherapy (RT) technology commonly adopted into the clinic. As a standard procedure, the patient-specific quality assurance (QA) will be performed prior to the RT delivery to evaluate the quality of the RT plan, which is recommended by the report of the American Association of Physicists in Medicine Task Group (TG) Nos. 119 and 218 (Ezzell et al., 2009; Miften et al., 2018). In the evaluation, a gamma passing rate (GPR) is derived from the measured and calculated dose of the treatment planning system (TPS). Under the recommended gamma action limit (Ezzell et al., 2009), the plan is determined as "pass" or "fail." However, the implementation of QA is time-consuming and adds a heavy clinical workload to the physicist.

With the development of technology, machine learning (ML) is widely applied in medical research (Jiang et al., 2021; Ma et al., 2021; Xia et al., 2021; Zhang et al., 2021a,b). To enhance the efficiency of QA, instead of the measured-based method, the ML technique has been involved in predicting the GPR by using the features of intensity-modulated radiotherapy (IMRT) or VMAT (Chiavassa et al., 2019; Chan et al., 2020; Kalet et al., 2020; Tomori et al., 2021). For example, Valdes et al. (2016) adopted the regression algorithm LASSO to predict the GPR of IMRT plans based on 78 aperture-based complexity metrics. In the following year, they investigated the prediction of GPR with different measurement techniques and across multiple institutions using the same algorithm (Valdes et al., 2017). All results showed that the ML technique is an efficient tool to predict the GPR. In addition to the prediction, Li et al. (2019) performed a multiple institutions analysis on the classification and prediction of GPR accounting for the VMAT plan by using ML algorithms and 54 complexity metrics. Ono et al. (2019) involved 28 complexity metrics to predict the VMAT GPR by building two models using multiple regression analysis and neural networks. All the above studies predict the plan GPR based on the complexity metric-based/aperture-based features (Du et al., 2014). The complexity metric is calculated by adopting the multiple leaf collimator (MLC) position, jaw position, gantry angle, and monitor units (MU) stored in the RT plan. A series of metrics (McNiven et al., 2010; Younge et al., 2012; Masi et al., 2013; Du et al., 2014; Park et al., 2014, 2015; Crowe et al., 2015; Götstedt et al., 2015; Sumida et al., 2017; Lam et al., 2019) have been developed to characterize the plan complexity correlated with the measured GPR, and those results demonstrate a good correlation with the measured GPR. Moreover, following the suggestion from Du et al. (2014), however, all complexity metrics attempt to characterize the treatment plan accuracy using a single complexity score, but they do not adequately distinguish plan heterogeneity.

Along with the plan complexity, another modeling approach, known as one of the fluence map complexity-based approaches, was developed. Nauta et al. (2011) analyzed the fluence map complexity employing fractal dimensions analysis. Interian et al. (2018) predicted the GPR of the IMRT plan using fluence maps-based features. Park et al. (2019) performed the texture analysis method on fluence maps of the VMAT plan. The fluence map for a VMAT plan is generated by superposing each fluence map of the control point. Tomori et al. (2018) adapted the dummy plans to train a model by using the two planar dose distributions. This method can make the modeling more efficient. To improve the model performance, Hirashima et al. (2020) combined complexity features and dosiomics features to predict and classify the GPR. However, the fluence-based approach has a limitation in that some fluence maps can be produced either by a single large beam or by a combination of successive small beams (Du et al., 2014).

The features of either the aperture-based approach or fluence map-based approach only consider the overall information of the treatment plan or dose distribution. For example, the complexity metrics of the modulation complexity score (MCS), small aperture score (SAS), plan area (PA), plan irregularity (PI), etc. are all overall characteristic metrics. Even with plenty of features extracted from the fluence map-based approach, the fluence map is generated from a superposition map. To overcome the limitation of using an overall feature, more attention should be paid to the information from each control point (Shiba et al., 2020). Inspired from the fluence map-based method

(Park et al., 2019), the aperture-based metrics at each control point can be calculated based on the RT plan DICOM file, and then a series of statistical histogram data using metrics of all control points, noted as planomics, can be extracted as plan features.

Overall, this study will: (1) propose a new feature extraction method by considering complexity metrics at each control point and thereby (2) investigate the performance of new feature-based prediction and classification models with ML by using the head and neck (H&N) VMAT plan of the patient; (3) additionally, the combination of the two features will also be used to predict the GPR.

MATERIALS AND METHODS

Data Characteristics

A total of 131 H&N plans in our institution from 2019 to 2021 were retrospectively collected. All plans adapted the 6 MV X-ray with VMAT technique by using 2–7 arcs, which were designed in the Eclipse^{TM} treatment planning system (V15.6, Varian Medical Systems, Palo Alto, CA, United States) using the medical linac of TrueBeam, VitalBeam, and Halcyon (Varian Medical Systems). These three models of medical linac are equipped with an electronic portal imaging device (AS1000). For TrueBeam and VitalBeam, 120 pairs of the MLC, whose maximum field size is 40 × 40 cm with 5 mm leaf width for central 20 cm of field and 10 mm for outer 20 cm of field. The Halcyon has a dual-layer MLC system with 57 pairs of MLC forming a maximum field size of 28 × 28 cm and 5 mm leaf width. In plan designing, a CT image with 3 mm slice spacing and grid spacing of the dose calculation with 3 × 3 × 3 mm was used to calculate dose distribution using algorithms of AAA version 15.6 in Eclipse. Besides, the dose rates of the plan and the minimum leaf gap were set as 600 MU/min (TrueBeam and VitalBeam) or 800 MU/min (Halcyon) and 0.2 cm, respectively. Moreover, all plans were optimized using gantry angle sampling of 2.0341° between the control points for Eclipse.

Dosimetric Verification

Dosimetric verification for all plans was carried out in the corresponding linac by using the portal dosimetry system (Version 15.6) integrated into Eclipse based on the EPID. A comparison between the measured dose and the TPS-calculated dose for all plans was performed using gamma analysis in terms of three gamma indices of 3%/3 mm, 3%/2 mm, and 2%/2 mm with a 10% dose threshold. Besides, the absolute dose model and global normalization were adapted to the gamma analysis.

Conventional Features

Forty-eight conventional features affecting the dose delivery accuracy were used in the study, as shown in **Table 1** (McNiven et al., 2010; Nauta et al., 2011; Younge et al., 2012; Masi et al., 2013; Du et al., 2014; Park et al., 2014, 2015; Crowe et al., 2015; Götstedt et al., 2015; Li et al., 2019). All features were calculated by an in-house-developed Python script based on the RT plan DICOM files extracted from the Eclipse TPS. The rest of the six features used in the paper (Li et al., 2019) were not included in our study owing to the non-consistent number of VMAT arcs used in the treatment plan. In addition, the angle range of the VMAT arc for some patient plans was also non-consistent. In this case, the Python script has been developed specifically to address those problems. Besides, the script has been updated to deal with the difference in dose rates and the number of MLC leaves.

Planomics Features

The planomics features were calculated using the DICOM-RT plan files. The planomics feature is a kind of control point-based complexity metric. The calculation and definition of the planomics features were integrated into **Supplementary Section 1**. All features were calculated using an in-house Python code.

TABLE 1 | Conventional plan feature metrics.

Quantity	Metrics	References
12	Modulation index for leaf speed (MI_s, $f = [0.2, 0.5, 1, 2]$)	Park et al., 2014
	Modulation index for leaf acceleration (MI_a, $f = [0.2, 0.5, 1, 2]$)	
	Modulation index for total (MI_t, $f = [0.2, 0.5, 1, 2]$)	
13	The proportion of leaf speed ranging from $a_i - a_{i1}$ ($S_{a_i - a_{i1}}$) with $a = [0, 0.4, 0.8, 1.2, 1.6, 2.0]$ and $i = 1, 2, \ldots, 5$.	Park et al., 2015
	The proportion of leaf acceleration ranging from $b_i - b_{i1}$ ($A_{b_i - b_{i1}}$) with $b = [0, 1, 2, 4, 6]$ and $i = 1, 2, 3, 4$.	
	Average leaf speed (ALS), standard deviation of leaf speed (SLS) Average leaf acceleration (ALA), standard deviation of leaf acceleration (SLA)	
4	Small aperture score (SAS, $l = 5, 10, 20$ mm) Mean asymmetry distance (MAD)	Crowe et al., 2015
3	Modulation complex score (MCS), leaf sequence variability (LSV), aperture area variability (AAV)	McNiven et al., 2010
5	Plan area (PA), plan irregularity (PI), plan modulation (PM), plan normalized MU (PMU), union aperture area (UAA)	Du et al., 2014
1	Edge metric (EM)	Younge et al., 2012
3	Converted aperture metric (CAM), edge area metric (EAM), circumference/area (C/A)	Götstedt et al., 2015
2	Average leaf travel distance (ALT), combination of ALT and MCS (ALTMCS)	Masi et al., 2013
2	Average leaf gap (ALG), standard deviation of leaf gap (SLG)	Nauta et al., 2011
3	Average dose rate (ADR), standard deviation of dose rate (SDR), prescribed dose to primary target per fraction (dose)	Li et al., 2019

Machine Learning Model Construction and Validation

The overall workflow of this study is shown in **Figure 1**. The DICOM plan data are firstly extracted from the TPS, and then conventional and planomics features were calculated based on the above approach. After combining with clinical outcome, two types of the model will be built, namely, the (a) prediction model and (b) classification model. Finally, the model was evaluated by using a multiple train–test score based on evaluation metrics. In the study, the models of regression prediction and classification are trained based on the Python package Scikit-learn (Pedregosa et al., 2011).

In the prediction model construction, the procedure of model construction is shown in **Figure 2A**. As shown in the figure, all feature data were divided into two parts: training (70%) and testing (30%) cohorts (train–test split). In the training cohort, the stable and predictive features were firstly screened out in the step of feature selection. Following that, a gradient boosting regressor (GBR) was adapted to construct the prediction model. To get an optimal regression model, two improved approaches were used: (1) random grid search for achieving optimal hyperparameter of GBR and (2) fivefold cross-validation for tuning parameters in the model.

In the feature selection (**Figure 3**), the features were: (1) firstly removed by using an unsupervised feature selection method of the analysis of variance (ANOVA) with zero thresholds, and then (2) a large number of features were selected based on the *F*-score by using the *F*-test-based univariate feature selection method. To obtain a set of features with stability and robustness, features were selected by repeating the above two steps for a hundred iterations. The 30 features with the highest frequency were screen out based on the number of occurrences in the 100 sets of selected features. Eventually, the *k* features were selected by using minimum redundancy maximum relevance (MRMR). It should be mentioned that the testing label in the 100 iterations is not involved in eliminating the effect of data leakage. Besides, the appropriated maximum number of features in a model should be smaller than 10% of the sample size (Abu-mostafa, 1995). Therefore, in the study, the *k* is set as 10 for building models.

In the classification model construction, the model procedure of model construction is shown in **Figure 2**. The model will be constructed by following almost the same procedure of the prediction model construction. One difference is that a classifier of linear least squares with l2 regularization (Ridge) was used to build the classification model. Another difference is that the whole modeling process will repeat 20 times to evaluate the overall predictability performance of the classification model using several different feature modalities. Based on the TG 218 report (Miften et al., 2018), the action limit was used to act VMAT plan as "pass" or "fail." In the study, stricter action limits were chosen with 99% for 3%/3 mm, 98% for 3%/2 mm, and 95% for 2%/2 mm.

According to the types of feature modality, three kinds of the model were constructed separately using the training cohort for the prediction and classification models, respectively: (1) using conventional features (CF), (2) using planomics features (PF), and (3) using hybrid features (HF). After feature modeling, a test cohort was used to assess the performance of two kinds of models by using one metric of the absolute prediction error (APE) using $APE = \left|GPR_{pred} - GPR_{meas}\right|/GPR_{meas}$ for the prediction model and one metric of the area under the receiver operating characteristic (ROC) curve (AUC) for the classification model.

RESULTS

Measured Gamma Passing Rates

Between 2019 and 2021, 131 valid patient plan data were involved in the study after data cleaning based on plan data integrity. The distributions of GPR for total, training, and test cohorts are listed in **Table 2**. As shown in **Table 2**, the majority of GPR was distributed into 95–100% for 3%/3 mm, 90–100% for 3%/2 mm, and 85–98% for 2%/2 mm. Based on the distributions, the strict gamma threshold and ratios of the corresponding samples were 99% and 53.9%/46.1% for 3%/3 mm, 98% and 48.4%/51.6% for 3%/2 mm, and 95% and 40.0%/60.0% for 2%/2 mm.

Feature Information

In the feature extraction, a total of 48 conventional features and 2,476 planomics features for each VMAT plan were calculated by analyzing the plan DICOM files using the package of Pydicom 2.1.2 based on Python 3.7 (Mason, 2011). After feature selection,

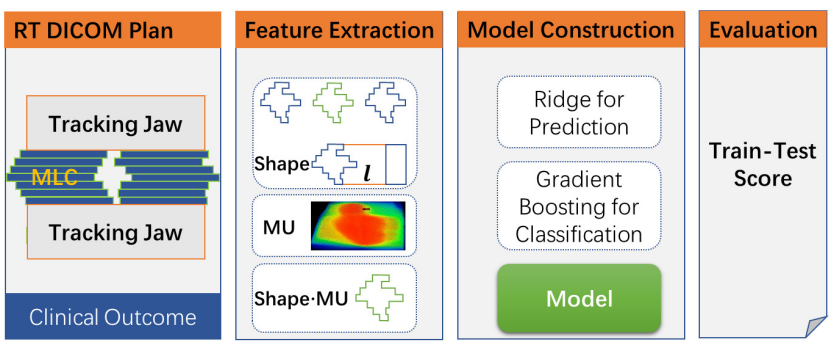


FIGURE 1 | The overall workflow for the study.

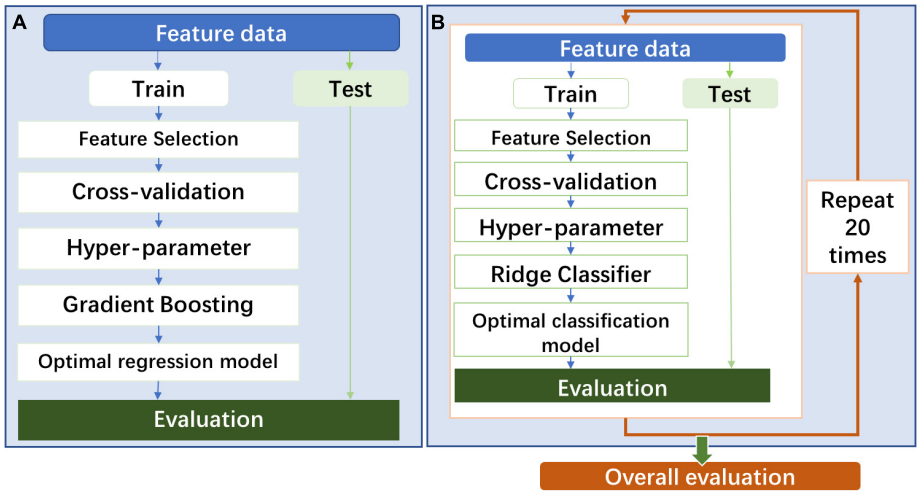


FIGURE 2 | (A) The procedure of prediction model construction. **(B)** The procedure of classification model construction.

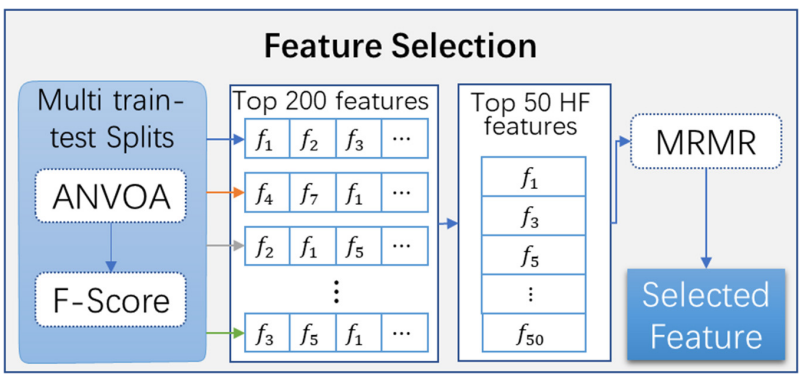


FIGURE 3 | The method for feature selection. HF, high frequency; ANOVA, analysis of variance; MRMR, minimum redundancy maximum relevance.

the used features in the prediction and classification models are collected in **Supplementary Tables 4, 5**.

Model Performance

The hyperparameters in the three regression models using the hybrid features are all the same with alpha of 0.12, learning rate of 0.12, number of estimators of 58, and maximum depth of 3.

TABLE 2 | The summary information of different gamma criteria.

Gamma indices	3%/3 mm	3%/2 mm	2%/2 mm
	Total, *N* (%)	Total, *N* (%)	Total, *N* (%)
[99%, 1]	86 (57.3)	51 (34.0)	22 (14.7)
[98%, 99%]	28 (18.7)	27 (18.0)	10 (6.7)
[95%, 98%]	33 (22.0)	43 (28.7)	35 (23.3)
[90%, 95%]	3 (2.0)	24 (16.0)	56 (37.3)
[85%, 90%]	0 (0)	5 (3.3)	18 (12.0)
[70%, 85%]	0 (0)	0 (0)	9 (6.0)
Mean ± SD	[98.7 ± 1.5] %	[97.3 ± 2.7] %	[93.8 ± 5.0] %

In the classification model, 30 classification models using hybrid features for each GPR were constructed, and the hyperparameter in the Ridge classifier was different model to model with the two most frequent alpha values of 31.6 and 1.0.

Prediction Accuracy

The results of prediction accuracy are shown in **Figure 4**. At gamma criteria of 2%/2 mm, the average APE (red plus) by using CF, PF, and HF was 1.3 ± 1.2%, 1.7 ± 1.5%, and 1.1 ± 1.0% in the training cohort and 3.6 ± 3.0%, 3.8 ± 3.5%, and 4.1 ± 3.1% in the testing cohort, respectively. At gamma criteria of 3%/2 mm, the average APE (red plus) by using CF, PF, and HF was 0.7 ± 0.6%, 1.0 ± 1.1%, and 0.6 ± 0.6% in the training cohort and 2.0 ± 2.0%, 2.2 ± 1.8%, and 2.2 ± 1.9% in the testing cohort, respectively. At gamma criteria of 3%/3 mm, the average APE (red plus) for using CF, PF, and HF was 0.4 ± 0.3%, 0.4 ± 0.5%, and 0.3 ± 0.3% in the training cohort and 1.2 ± 1.2%, 1.3 ± 1.0%, and 1.2 ± 1.1% in the testing cohort, respectively. All average APEs were smaller than 2.0 and 5.0% for the training and testing cohorts, respectively.

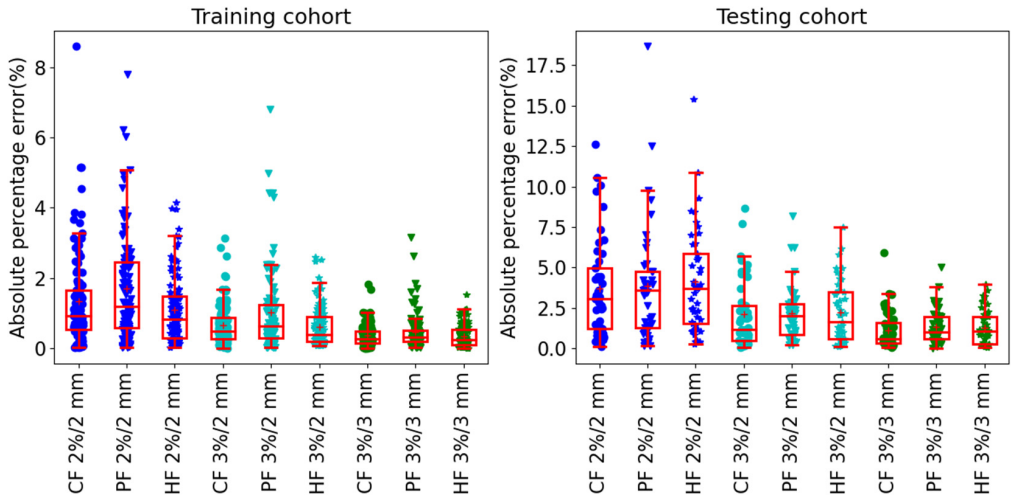


FIGURE 4 | The absolute prediction error by using different feature modalities at gamma criteria of 2%/2 mm, 3%/2 mm, and 3%/3 mm for the training **(left figure)** and testing **(right figure)** cohort, respectively. All solid circles, triangles, and pentagrams are the APE values. The red box shows the quartile values of each set of APE data. The red solid line and red plus inside the box are the median and average values, respectively.

Table 3 shows the summary of the average prediction error at three gamma criteria of 3%/3 mm, 3%/2 mm, and 2%/2 mm. There are no statistically significant prediction errors among each three prediction models at the three gamma criteria. More than 90% of the plans can be predicted precisely with lower than 3.0% APE for 3%/3 mm, 5.0% APE for 3%/2 mm, and 10% for 2%/2 mm. The MAPE of the three prediction models was smaller than 1.3% for 3%/3 mm, 2.2% for 3%/2 mm, and 4.1% for 2%/2 mm.

In all the three hybrid prediction models, the final selected features are the same, that is, *MCS*, $S_{0-0.4}$, $S_{0.4-0.8}$, $MI_{a\ 2-4}$, $MI_{a\ 4-6}$, *ALT*, *AAV*, *ALTMCS*, $MUAP_{TR\ 24,200-24,400}$, and $MUAP_{T\ 24,200-24,400}$. Eight features are from the conventional feature and the other two features are from the planomics feature.

Classification Accuracy

The classification results of the model with rigorous action limit (99, 98, and 95% for 3%/3 mm, 3%/2 mm, and 2%/2 mm, respectively) are shown in **Figure 5**. For the gamma criteria of 3%/3 mm, the average and SD AUCs of the training and testing cohorts were 0.67 ± 0.03/0.66 ± 0.07, 0.77 ± 0.03/0.73 ± 0.06, and 0.78 ± 0.02/0.75 ± 0.04, respectively. For the gamma criteria of 3%/2 mm, the average AUCs of the training and testing cohorts were 0.64 ± 0.03/0.62 ± 0.07, 0.70 ± 0.03/0.67 ± 0.06, and 0.75 ± 0.03/0.71 ± 0.07, respectively. For the gamma criteria of 2%/2 mm, the average AUCs of the training and testing cohorts were 0.72 ± 0.03/0.72 ± 0.06, 0.78 ± 0.04/0.73 ± 0.07, and 0.81 ± 0.03/0.75 ± 0.06, respectively.

From **Figure 5**, it is clear that the three hybrid classification models have the best performance in all the three gamma criteria. It is improved by 16 and 14% for the training and testing cohorts at 3%/3 mm, 17% and 11% for the training and testing cohorts at 3%/2 mm, and 13% and 4% for the training and testing cohorts at 2%/2 mm. Besides, the classification model using the planomics feature has better predictability compared with the conventional model. The final selected features in the three hybrid models are shown in **Supplementary Table 1**. Most of the features used in the hybrid model are planomics features, especially at 3%/2 mm with all planomics features.

TABLE 3 | The summary average prediction error in the testing cohort by using the three feature modalities at different gamma criteria.

Metrics		APE < 3.0%	APE < 5%	APE < 10%	MAPE (SD%)
3%/3 mm	CF, *n* (%)	42 (93.3)	44 (97.8)	45 (100)	1.2% (1.2)
	PF, *n* (%)	43 (95.6)	44 (97.8)	45 (100)	1.3% (1.0)
	HF, *n* (%)	41 (91.1)	45 (100)	45 (100)	1.2% (1.1)
3%/2 mm	CF, *n* (%)	35 (77.8)	42 (93.3)	45 (100)	2.0% (2.0)
	PF, *n* (%)	36 (80.0)	42 (93.3)	45 100)	2.1% (1.7)
	HF, *n* (%)	29 (64.4)	41 (91.1)	45 (100)	2.2% (1.8)
2%/2 mm	CF, *n* (%)	23 (51.1)	35 (77.8)	43 (95.6)	3.6% (3.1)
	PF, *n* (%)	19 (42.2)	35 (77.8)	43 (95.6)	3.8% (3.5)
	HF, *n* (%)	20 (44.4)	30 (66.7)	43 (95.6)	4.1% (3.2)

MAPE, mean absolute prediction error, SD, standard deviation; CF, conventional features; PF, planomics features; HF, hybrid features.

DISCUSSION

In this study, we proposed a novelty feature, namely, the planomics feature, to comprehensively characterize the quality of the radiotherapy plan with regression and classification analysis by using machine learning techniques. To the best of our knowledge, the features from each control point are firstly involved in the analysis of GPR. Only a similar study of Park et al. (2015) counted the MLC leaf speed and accelerations with a limit of five ranges for the VMAT plan, which had a limitation on the description of plan complexity. The other studies just

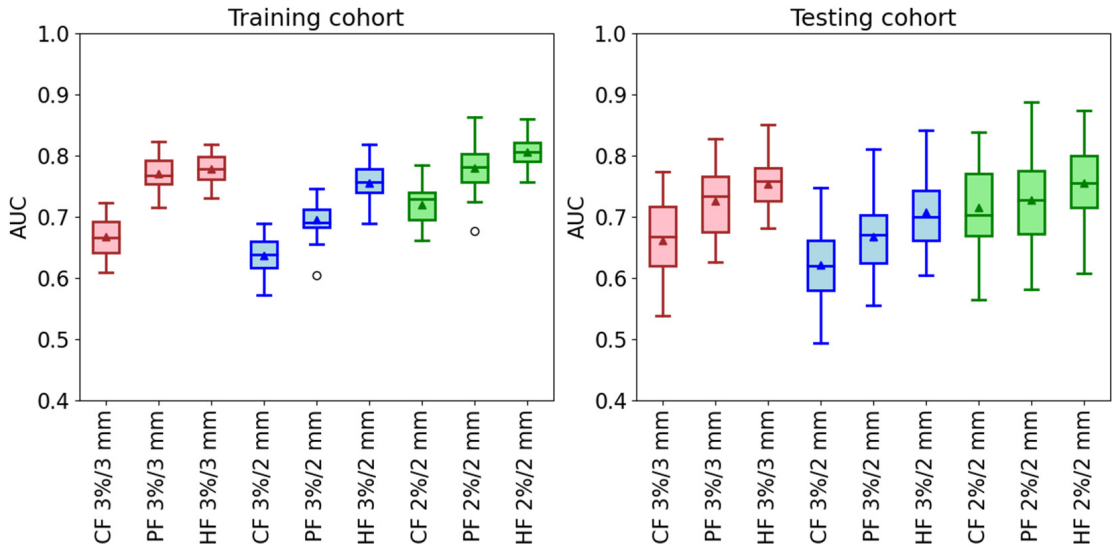


FIGURE 5 | The AUC of the classification model in the training and testing cohorts by using three kinds of feature modalities in the conditions of the three gamma criteria. The used action limit is 99% for 3%/3 mm, 98% for 3%/2 mm, and 95% for 2%/2 mm.

adapted overall values calculated from each control point. For example, McNiven et al. (2010) adapted the summation to calculate MCS, AAV, and LSV. In the study of Du et al. (2014), they proposed features (i.e., aperture area, aperture shape irregularity, and degree of beam modulation) that were calculated by the summation of each control point value. Likewise, Younge et al. (2012) applied the edge metric which was a summation using edge length and the aperture at each control point. In addition, Götstedt et al. (2015) utilized an average value of three metrics (convert edge metric, edge area metric, and the ratio *circumference/area*) at each point. The other complexity metrics in **Table 1** are calculated based on the methods of summation, average, or standard deviation.

For the model performance, the regression results show that the APE has consistent results for all prediction models at the three gamma criteria, which agreed with the result of the study of Li et al. (2019) that more than 90% of plans had APE <3.5% for 2%/2 mm and 5% for 3%/2 mm. In the classification, the results demonstrated well that our proposed planomics features can distinguish heterogeneity in terms of the H&N VMAT plan more efficiently compared with the conventional features. Besides, hybrid models for three GPR thresholds achieve the best performance with the AUC increment of 4–16% in the testing cohort compared with the conventional and planomics models.

In the three hybrid regression models, only two MUAP planomics features were selected. It shows that the conventional features have stronger correlations with the measured GPR. In contrast, for the hybrid classification model, the selected features are dominated by the planomics features. It means that the planomics have a better relevance with GPR results of "pass" or "fail." Besides, for the selected features (see **Supplementary Table 5**), most of the planomics features are related to MU or MUAP with seven, eight, and four features for 3%/3 mm, 3%/2 mm, and 2%/2 mm, respectively. It indicates that more attention should be paid to the MU in the planning design. One PSR feature, similar to the BI metric (Du et al., 2014), is contained in all three hybrid classification models. It shows a good correlation with the GPR results for different gamma criteria. In addition, SF features are included in the hybrid classification model of 3%/3 mm and 3%/2 mm. Only two RF features and one area feature are used in the classification model at 2%/2 mm, which have a weaker connection with GRP results.

Based on the recommendation of the TG 218 report (Ezzell et al., 2009), the action limit of 3%/2 mm is 90% with a 10% dose threshold. In our study, stricter action limits were used with 98% at 3%/2 mm with a 10% dose threshold. As mentioned by Li et al. (2019), to achieve an adequate amount of low GPR plans for modeling is too difficult to build a balanced dataset in a single institution. The imbalanced dataset is common in medical diagnostics, text classification, face, and image recognition, and also in GPR studies (Valdes et al., 2016; Tomori et al., 2018; Li et al., 2019). The imbalanced data can affect classifiers to weaken the performance of the classification model, as most classifier learning algorithms are accountable for a relatively balanced distribution (Sun et al., 2009; Kaur et al., 2019; Thabtah et al., 2020). Solutions on the imbalance are to obtain more data or adopt a specific algorithm (Kaur et al., 2019). To avoid the effect from the imbalanced data, a straightforward way is to set stricter action limits to obtain a relatively balanced class data distribution. Based on the data distribution, the stricter action limits of 99, 98, and 95% were chosen for 3%/3 mm, 3%/2 mm, and 2%/2 mm, respectively. In addition, another set of action limit with 98, 95, and 90% was chosen to evaluate the classification model performance in all gamma criteria. The modeling results using multiple train–test splits are shown in **Supplementary Figure 3**. The data distribution with the label "pass" and "fail" was 76.0%/24.0%

at 3%/3 mm, 80.7%/19.3% at 3%/2 mm, and 82.0%/18.0% at 2%/2 mm. In the modeling, a downsampling method was used to reduce the effect of imbalanced data. The figure presents the same results with **Figure 5**, where our proposed planomics features have better predictability and the hybrid models give the best results, too.

Instead of a single train–test split, an approach with multiple train–test splits was used to evaluate the predictability of three sets of features to classify the GPR result. One main concern is that a single train–test split generates a random feature data distribution in training and testing cohorts. In some cases, one set of features has a stronger correlation in two cohorts, yet another set of features has a weaker correlation in two cohorts. In other cases, the result may give an opposite result. To avoid this randomness, the approach with multiple train–test splits was involved in the study.

The shortcoming of the study is that only H&N VMAT plans were involved, and data for just only one institution were included in the study. For the VMAT plans of the other site, the IMRT plans, and data from multiple institutions, the performance of the planomics should be further investigated. Besides, the other complexity metrics will also be evaluated in a future work.

CONCLUSION

Our proposed control point-based planomics feature can be used to predict and classify the measured GPR for patient-specific quality assurance. In the regression prediction, the planomics and conventional features give a similar modeling performance. In the classification, the predictability of the planomics feature is better than the conventional aperture-based complexity metric. Besides, the combination using planomics and conventional features provides the best result in the classification.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The study was reviewed and approved by the hospital ethical committee of the Affiliated Cancer Hospital of Zhengzhou University.

AUTHOR CONTRIBUTIONS

BL and JC equally contribute to the study for data analyzing and writing. HG designed and organized the study. WG, RM, and HL collected the image data. ZL, XC, TC, and DL carried out the dose measurement of patient QA. XZ, TW, and HT collected the clinical information. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (Nos. 81773230 and 82002469) and the Scientific Research Foundation for Doctoral of the Affiliated Cancer Hospital of Zhengzhou University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.744296/full#supplementary-material

REFERENCES

Abu-mostafa, Y. S. (1995). Hints. *Neural Comput.* 7, 639–671.

Chan, M. F., Witztum, A., and Valdes, G. (2020). Integration of AI and machine learning in radiotherapy QA. *Front. Artif. Intel.* 3:577620. doi: 10.3389/frai.2020.577620

Chiavassa, S. I, Bessieres, M. Edouard, Mathot, M., and Moignier, A. (2019). Complexity metrics for IMRT and VMAT Plans: a review of current literature and applications. *Br. J. Radiol.* 92:20190270. doi: 10.1259/bjr.20190270

Crowe, S. B., Kairn, T., Middlebrook, N., Sutherland, B., Hill, B., Kenny, J., et al. (2015). Examination of the properties of IMRT and VMAT beams and evaluation against pre-treatment quality assurance results. *Phys. Med. Biol.* 60, 2587–2601. doi: 10.1088/0031-9155/60/6/2587

Du, W., Cho, S. H., Zhang, X., Hoffman, K. E., and Kudchadker, R. J. (2014). Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. *Med. Phys.* 41:021716. doi: 10.1118/1.4861821

Ezzell, G. A., Burmeister, J. W., Dogan, N., Losasso, T. J., Mechalakos, J. G., Mihailidis, D., et al. (2009). IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. *Med. Phys.* 36, 5359–5373. doi: 10.1118/1.3238104

Götstedt, J., Hauer, A. K., and Bäck, A. (2015). Development and Evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings. *Med. Phys.* 42, 3911–3921. doi: 10.1118/1.4921733

Hirashima, H., Ono, T., Nakamura, M., Miyabe, Y., Mukumoto, N., Iramina, H., et al. (2020). Improvement of prediction and classification performance for gamma passing rate by using plan complexity and dosiomics features. *Radiother. Oncol.* 153, 250–257. doi: 10.1016/j.radonc.2020.07.031

Interian, Y., Rideout, V., Kearney, V. P., Gennatas, E., Morin, O., Cheung, J., et al. (2018). Deep nets vs expert designed features in medical physics: an IMRT QA case study. *Med. Phys.* 45, 2672–2680. doi: 10.1002/mp.12890

Jiang, Y., Zhang, Y., Lin, C., Wu, D., and Lin, C. T. (2021). EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system. *IEEE Trans. Intel. Transport. Syst.* 22, 1752–1764. doi: 10.1109/TITS.2020.2973673

Kalet, A. M., Luk, S. M. H., and Phillips, M. H. (2020). Radiation therapy quality assurance tasks and tools: the many roles of machine learning. *Med. Phys.* 47, e168–e177. doi: 10.1002/mp.13445

Kaur, H., Pannu, H. S., and Malhi, A. K. (2019). A systematic review on imbalanced data challenges in machine learning: applications and solutions. *ACM Comput. Surv.* 52, 1–36. doi: 10.1145/3343440

Lam, D., Zhang, X., Li, H., Deshan, Y., Schott, B., Zhao, T., et al. (2019). Predicting gamma passing rates for portal dosimetry-based IMRT QA using machine learning. *Med. Phys.* 46, 4666–4675. doi: 10.1002/mp.13752

Li, J., Wang, L., Zhang, X., Liu, L., Li, J., Chan, M. F., et al. (2019). Machine learning for patient-specific quality assurance of VMAT: prediction and classification

accuracy. *Int. J. Rad. Oncol. Biol. Phys.* 105, 893–902. doi: 10.1016/j.ijrobp.2019.07.049

Ma, C., Wang, R., Zhou, S., Wang, M., Yue, H., Zhang, Y., et al. (2021). The structural similarity index for IMRT quality assurance: radiomics-based error classification. *Med. Phys.* 48, 80–93. doi: 10.1002/mp.14559

Masi, L., Doro, R., Favuzza, V., Cipressi, S., and Livi, L. (2013). Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy. *Med. Phys.* 40:071718. doi: 10.1118/1.4810969

Mason, D. (2011). SU-E-T-33: pydicom: an open source DICOM library. *Med. Phys.* 38, 3493–3493.

McNiven, A. L., Sharpe, M. B., and Purdie, T. G. (2010). A new metric for assessing IMRT modulation complexity and plan deliverability. *Med. Phys.* 37, 505–515. doi: 10.1118/1.3276775

Miften, M., Olch, A., Mihailidis, D., Moran, J., Pawlicki, T., Molineu, A., et al. (2018). Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM task group No. 218. *Med. Phys.* 45, e53–e83. doi: 10.1002/mp.12810

Nauta, M., Villarreal-Barajas, J. E., and Tambasco, M. (2011). Fractal analysis for assessing the level of modulation of IMRT fields. *Med. Phys.* 38, 5385–5393. doi: 10.1118/1.3633912

Ono, T., Hirashima, H., Iramina, H., Mukumoto, N., Miyabe, Y., Nakamura, M., et al. (2019). Prediction of dosimetric accuracy for VMAT plans using plan complexity parameters via machine learning. *Med. Phys.* 46, 3823–3832. doi: 10.1002/mp.13669

Park, J. M., Kim, J. I., and Park, S. Y. (2019). Prediction of VMAT delivery accuracy with textural features calculated from fluence maps. *Rad. Oncol.* 14:235. doi: 10.1186/s13014-019-1441-7

Park, J. M., Park, S. Y., Kim, H., Kim, J. H., Carlson, J., and Ye, S. J. (2014). Modulation indices for volumetric modulated arc therapy. *Phys. Med. Biol.* 59, 7315–7340. doi: 10.1088/0031-9155/59/23/7315

Park, J. M., Wu, H.-G., Kim, J. H., Carlson, J. N. K., and Kim, K. (2015). The effect of MLC speed and acceleration on the plan delivery accuracy of VMAT. *Br. J. Radiol.* 88:20140698. doi: 10.1259/bjr.20140698

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python fabian. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.1289/EHP4713

Shiba, E., Saito, A., Furumi, M., Kawahara, D., Miki, K., Murakami, Y., et al. (2020). Predictive gamma passing rate for three-dimensional dose verification with finite detector elements via improved dose uncertainty potential accumulation model. *Med. Phys.* 47, 1349–1356. doi: 10.1002/mp.13985

Sumida, I., Yamaguchi, H., Das, I. J., Kizaki, H., Aboshi, K., Tsujii, M., et al. (2017). Organ-specific modulation complexity score for the evaluation of dose delivery. *J. Rad. Res.* 58, 675–684. doi: 10.1093/jrr/rrw129

Sun, Y., Wong, A. K. C., and Kamel, M. S. (2009). Classification of imbalanced data: a review. *Int. J. Pattern Recogn. Artif. Intel.* 23, 687–719. doi: 10.1142/S0218001409007326

Thabtah, F., Hammoud, S., Kamalov, F., and Gonsalves, A. (2020). Data imbalance in classification: experimental evaluation. *Inform. Sci.* 513, 429–441. doi: 10.1016/j.ins.2019.11.004

Tomori, S., Kadoya, N., Kajikawa, T., Kimura, Y., Narazaki, K., Ochi, T., et al. (2021). Systematic method for a deep learning-based prediction model for gamma evaluation in patient-specific quality assurance of volumetric modulated arc therapy. *Med. Phys.* 48, 1003–1018. doi: 10.1002/mp.14682

Tomori, S., Kadoya, N., Takayama, Y., Kajikawa, T., Shima, K., Narazaki, K., et al. (2018). A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance. *Med. Phys.* 45, 4055–4065. doi: 10.1002/mp.13112

Valdes, G., Chan, M. F., Lim, S. B., Scheuermann, R., Deasy, J. O., and Solberg, T. D. (2017). IMRT QA using machine learning: a multi-institutional validation. *J. Appl. Clin. Med. Phys.* 18, 279–284. doi: 10.1002/acm2.12161

Valdes, G., Scheuermann, R., Hung, C. Y., Olszanski, A., Bellerive, M., and Solberg, T. D. (2016). A mathematical framework for virtual IMRT QA using machine learning. *Med. Phys.* 43, 4323–4334. doi: 10.1118/1.4953835

Xia, K., Zhang, Y., Jiang, Y., Qian, P., Dong, J., Yin, H., et al. (2021). TSK fuzzy system for multi-view data discovery underlying label relaxation and cross-rule & cross-view sparsity regularizations. *IEEE Trans. Industr. Inf.* 17, 3282–3291. doi: 10.1109/TII.2020.3007174

Younge, K. C., Matuszak, M. M., Moran, J. M., McShan, D. L., Fraass, B. A., and Roberts, D. A. (2012). Penalization of aperture complexity in inversely planned volumetric modulated arc therapy. *Med. Phys.* 39, 7160–7170. doi: 10.1118/1.4762566

Zhang, Y., Wang, G., Chung, F., and Wang, S. (2021a). Support vector machines with the known feature-evolution priors. *Knowledge Based Syst.* 223:107048. doi: 10.1016/J.KNOSYS.2021.107048

Zhang, Y., Wang, S., Xia, K., Jiang, Y., and Qian, P. (2021b). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inf. Fus.* 66:170.

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

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

Copyright © 2021 Li, Chen, Guo, Mao, Zheng, Cheng, Cui, Lou, Wang, Li, Tao, Lei and Ge. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 05 October 2021
doi: 10.3389/fnins.2021.749232

Evolutionary Multitasking-Based Multiobjective Optimization Algorithm for Channel Selection in Hybrid Brain Computer Interfacing Systems

Tianyu Liu[1], *Zhixiong Xu*[1], *Lei Cao*[1]* *and Guowei Tan*[2]*

[1] School of Information Engineering, Shanghai Maritime University, Shanghai, China, [2] Department of Neurosurgery, Xiamen Key Laboratory of Brain Center, The First Affiliated Hospital of Xiamen University, Xiamen, China

Hybrid-modality brain-computer Interfaces (BCIs), which combine motor imagery (MI) bio-signals and steady-state visual evoked potentials (SSVEPs), has attracted wide attention in the research field of neural engineering. The number of channels should be as small as possible for real-life applications. However, most of recent works about channel selection only focus on either the performance of classification task or the effectiveness of device control. Few works conduct channel selection for MI and SSVEP classification tasks simultaneously. In this paper, a multitasking-based multiobjective evolutionary algorithm (EMMOA) was proposed to select appropriate channels for these two classification tasks at the same time. Moreover, a two-stage framework was introduced to balance the number of selected channels and the classification accuracy in the proposed algorithm. The experimental results verified the feasibility of multiobjective optimization methodology for channel selection of hybrid BCI tasks.

Keywords: brain-computer interfaces, channel selection, evolutionary multitasking, multiobjective optimization, two-stage framework

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jiaqi Zhao,
China University of Mining and Technology, China
Lingling Li,
Xidian University, China

***Correspondence:**
Lei Cao
lcao@shmtu.edu.cn
Guowei Tan
gwtanxm@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *29 July 2021*
Accepted: *08 September 2021*
Published: *05 October 2021*

Citation:
Liu T, Xu Z, Cao L and Tan G (2021) Evolutionary Multitasking-Based Multiobjective Optimization Algorithm for Channel Selection in Hybrid Brain Computer Interfacing Systems. Front. Neurosci. 15:749232. doi: 10.3389/fnins.2021.749232

1. INTRODUCTION

Brain-computer interface (BCI) develops a communication means between brains and external devices (Alcaide-Aguirre and Huggins, 2014). This technology is used for developing assistive devices for helping paralyzed patients or control gaming. Due to its convenience, electroencephalography (EEG) is widely adopted in non-invasive BCIs for multi-channel signal acquisition (Kevric and Subasi, 2017). In general, more channels would provide more information that can improve the performance of BCI classification. However, channel selection is essential to filter redundant information and simplify experimental manipulation. For example, MI tasks may only need a number of channels as few as 3–5 without lowering the classification accuracy according to temporal-spatio analysis (Pfurtscheller and da Silva, 1999). Furthermore, using more channels is inconvenience for clinical operation and signal processing. Therefore, how to make a compromise between channel number and classification accuracy is a worth-probing question, especially for hybrid BCIs in real-world applications.

In previous works, several effective methods have been proposed for channel selection, such as filter-based methods (Yang et al., 2016), wrapper-based methods (Qiu et al., 2016), correlation-based methods (Jin et al., 2019), machine learning methods (Su et al., 2019), Heuristic searching

methods (Sun et al., 2021), and so on. As mentioned before, channel selection needs to make a compromise between two objectives: the number of selected channels and classification accuracy. This perfectly coincides with the goal of multiobjective evolutionary algorithms (MOEAs), which is to make a compromise among multiple optimization objectives. In recent years, multiobjective evolutionary algorithms (MOEAs) have demonstrated their effectiveness in channel selection. For example, the classic Non-dominated Sorting Genetic Algorithm-II (NSGA-II), Multiobjective Evolutionary Algorithm Based on Decomposition (MOEA/D), and Multiobjective Particle Swarm Optimization (MOPSO) have been successfully applied for channel selection in the task of single-modality -based BCIs (Hasan et al., 2010; Moubayed et al., 2010; Kee et al., 2015). Current studies show that hybrid BCI, which combines motor imagery (MI) signals and steady-state visual evoked potentials (SSVEPs), can achieve certain goals more efficiently than single-modality -based BCI systems (Long et al., 2012; Ko et al., 2014). However, relevant works only focus on either the MI classification task (Jin et al., 2020; Sun et al., 2021) or the SSVEP classification task (Zhang et al., 2017; Ravi et al., 2019). Few works conduct channel selection for hybrid BCI tasks.

In this paper, a multitasking-based multiobjective evolutionary algorithm (EMMOA) is proposed to perform channel selection for MI and SSVEP tasks at the same time. In EMMOA, the problems of channel selection for both MI and SSVEP tasks can be regarded as multi-objective optimization problems (MOPs). In order to make a balance between classification accuracy and channel number, the problem of channel selection for MI task can be formulated as a two-objective optimization problem, which contains two objectives: the classification accuracy for MI task and the number of selected channels. Evolutionary multitasking mechanism, which is inspired by bio-cultural models of multifactorial inheritance (Gupta et al., 2015, 2016), provides the investigation of solving multiple tasks via a single population concurrently. In the evolutionary multitasking mechanism, different tasks will experience information transfer during the evolution process since they use the same population. Therefore, if multiple tasks are related then the searching process of solving one task may offer help in solving the other tasks (Gong et al., 2019; Bai et al., 2021). As the evolutionary multitasking mechanism is proved to be an efficient way to optimize multiple tasks, it has been introduced in EMMOA to optimize MI and SSVEP tasks simultaneously in this paper. Furthermore, EMMOA adopts a two-stage framework to improve searching efficiency. The first stage is based on an evolutionary multitasking mechanism and aims to obtain the Pareto-optimal solutions (PS) for MI and SSVEP tasks by one single population. The second stage is local searching. The second stage constructs a three-objective optimization problem, which used classification accuracy for MI task, classification accuracy for SSVEP task, and the number of selected channels as the optimization objectives. The constructed three-objective problem is optimized according to decision variable analysis based on the results obtained by the first stage.

The rest of this paper is organized as follows: Section 2 elaborates the data set, pre-processing and optimization methods used in this paper. Section 3 presents experimental and result analysis, followed by further discussion of this paper.

2. MATERIALS AND METHODS

2.1. Subjects and Data Acquisition

As shown in **Figure 1**, fifteen electrodes (i.e., "FC3," "FC4," "C5," "C3," "C1," "Cz," "C2," "C4," "C6," "CP3," "CP4," "POz," "O1," "Oz," and "O2") were used to acquire EEG signals by using a high-performance bio-signal amplifier. These electrodes were placed at the frontal, central, parietal and occipital regions. And the impedances were kept below 5 kΩ. EEG data was sampled at a frequency of 256 Hz and band pass filtered at 0.1–30 Hz.

In this experiment, 7 healthy volunteers, aged from 21 to 30, were selected for data acquisition. Subjects had no prior experience with the operation of hybrid BCI. They all gave informed consent approved by the Ethics Committee.

2.2. Feature Selection and Classification

In our work, MI tasks including left- and right-hand imagery are performed for pattern recognition. EEG signals are band pass filtered from 5 to 30 Hz firstly. Then the features are extracted by a well-known common spatial pattern (CSP) algorithm. The CSP method is useful for discriminating two populations of EEG dataset. The detailed methodology can be reviewed in Muller-Gerking et al. (2004). And a radial basis function kernel support vector machine (RBF-SVM) classifier is used for training these feature vectors.

From Lin et al. (2006), we utilize the canonical correlation analysis (CCA) method for SSVEP detection. CCA reflected the correlation relationship between EEG response signals and classical Fourier series at the stimulus frequency and its harmonics. It is widely used for spectral analysis in the biosignal process.

2.3. Multi-Objective Channel Selection Problem Formulation

Suppose the number of total channels is K, then a solution for the channel selection problem can be defined as a K-dimensional vector x, as shown in Equation (1). $x_i (1 \leq i \leq K)$ is called a decision variable. $x_i = 0$ means the corresponding channel is not selected. Otherwise, the corresponding channel is selected.

$$x = [x_1, x_2, ..., x_K], x_i \in \{0, 1\} \tag{1}$$

The purpose of channel selection problems is to select a smaller set of channels with as little sacrifice as possible in classification accuracy. For the MI classification task, the first objective function is the MI classification accuracy rate (MAR). The second objective function is defined as $NC = K - C$, where C is the number of selected channels (i.e., the number of elements whose value is 1 in solution x). Therefore, the ideal situation for an optimization algorithm is obtaining an optimal solution x^* that has the maximum values for both MAR and NC. However, it is impossible to get the above-mentioned x^*, since the confliction between MAR and NC. Thus, evolutionary optimization algorithms will obtain a Pareto Set

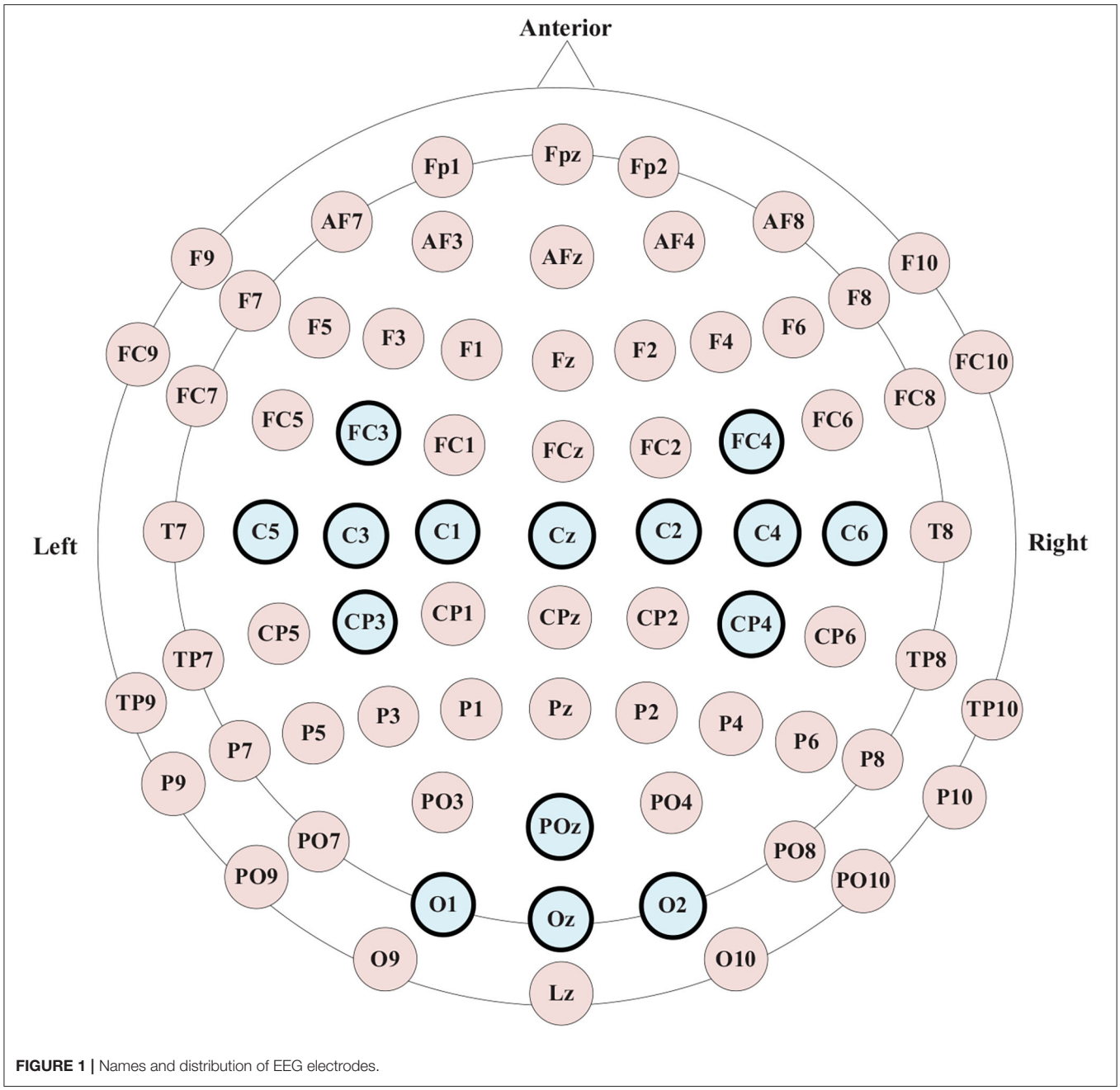


FIGURE 1 | Names and distribution of EEG electrodes.

(PS), which contains a set of Pareto-optimal solutions (non-dominated solutions) (Wang et al., 2018). One solution x is called a Pareto-optimal solution, if and only if there not exist any other solutions that are better than x for all objective functions. The Pareto optimal solutions are incomparable, since solution x may be better than solution y for one objective function but worse than y for the other one.

Similarly, the two objectives for SSVEP classification task are the SSVEP classification accuracy rate (*SAR*) and *NC*. In the proposed algorithm, the second stage aims to improve the optimization performance for both MI and SSVEP tasks by local searching, therefore a three-objective optimization problem, which uses *MAR*, *SAR*, and *NC* as the objective functions, is adopted in this stage.

2.4. Evolutionary Multitasking-Based Multiobjective Optimization Algorithm (EMMOA)

The framework of EMMOA is given in **Figure 2**. As shown in **Figure 2**, EMMOA adopts a two-stage framework. In EMMOA, the first stage is designed according to evolutionary multitasking mechanism and uses one single population to optimize two tasks (MI task and SSVEP task) simultaneously. In this case, the two

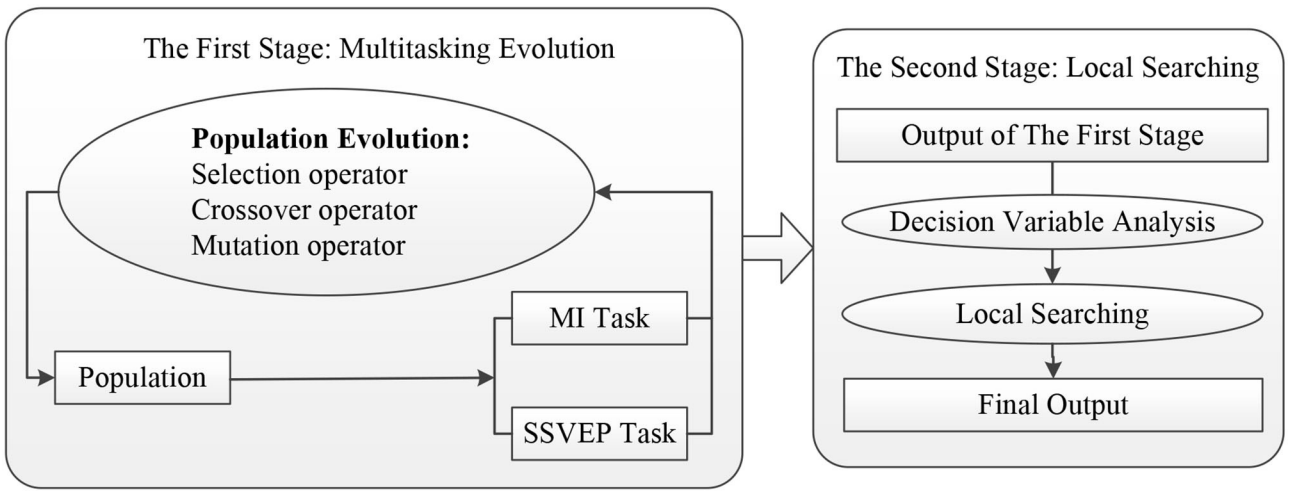


FIGURE 2 | Framework of EMMOA.

tasks will experience information transfer during the evolution process since they use the same population. The information transfer can share underlying similarities between the two tasks thereby facilitating improved optimization performance for both MI and SSVEP tasks. Based on the output of the first stage, the second stage aims to obtain the final output of the algorithm. With the help of decision variables analysis, the local searching strategy in the second stage makes a better compromise among MI classification accuracy, SSVEP classification accuracy, and the number of selected channels. More specifically, in the second stage, a decision variable analysis operator will be carried out on the Pareto-optimal sets obtained by the first stage and aims to divide the decision variables into different groups. The local searching operator will determine the searching directions of individuals according to the type of each decision variable and help the algorithm search more efficiently. The detailed description of the two stages is given below.

In the first stage, suppose the population size is N and the total number of channels is K, then the individual population can be initialized by generating N individuals (solutions) and each individual contains K elements. In this case, the individual population can be initialized as a $N \times K$ matrix. Each element in the individual population is uniformly generated from [0, 1]. If the value of an element is larger than 0.5, then the element will take the value of 1. Otherwise, the element will be set to 0. By associating each individual with a task label by the task assignment operator. In the initialization step, the task assignment operator is carried out by giving each individual the label, which is generated from $\{1, 2\}$ randomly. In the evolution process, the task labels of individuals are inherited from their parents. Each individual will optimize the task that is specified by its task label. For a task label, 1 and 2 represent the corresponding individual is assigned to optimize MI task and SSVEP task, respectively. The individual population is updated by using the population evolution step, which includes the tournament selection operator (Zhang et al., 2016), partial-mapped crossover operator (Ismkhan and Zamanifar, 2015), and flip-bit mutation operator (Chicano et al., 2015). The tournament selection operator, partial-mapped crossover operator, and simple mutation operator are classic operators, which are widely adopted in evolutionary algorithms to generate offspring populations (lines 1–21 in **Algorithm 1**). The task labels of offspring individuals are inherited from their parents in the evolution process (line 7 and line 19 in **Algorithm 1**). After updating the individual population (lines 22–23 in **Algorithm 1**), the Pareto-optimal sets for MI and SSVEP tasks will be updated accordingly (line 24 in **Algorithm 1**). Therefore, the first stage will output two Pareto-optimal sets (namely PS_MI and PS_SSVEP) for MI and SSVEP tasks separately. PS_MI contains a set of non-dominated solutions which aim to optimize *MAR* and *NC*, while PS_SSVEP contains a set of non-dominated solutions which aim to optimize *SAR* and *NC*. In this paper, the maximum size of both PS_MI and PS_SSVEP is set to 100. Take PS_MI for an example, if the size of PS_MI exceeds 100, then the non-dominated solutions in PS_MI will be decreased according to Crowding Distance (Deb et al., 2002). Crowding Distance, which is proposed to estimate the density of solutions, is widely adopted to maintain population diversity in EAs. The solution with a larger Crowding Distance value can be considered to make great contributions for maintaining population diversity. Therefore, if the number of the solutions in PS_MI exceeds 100, then the members with larger Crowding Distance values will be remain.

The second stage is implemented based on the output of the first stage, i.e., the current Pareto-optimal sets for MI and SSVEP tasks (PS_MI and PS_SSVEP). The purpose of the second stage is to obtain a set of Pareto-optimal solutions that make a compromise among MI classification accuracy, SSVEP classification accuracy, and the number of selected channels. Therefore, in this stage, the optimization problem becomes a maximum three-objective optimization with *MAR*, *SAR*, and *NC* as its objective functions. In this case, the Pareto-optimal

Algorithm 1 First stage of EMMOA.

Input:
POP (the individual population)
PS_MI and PS_SSVEP (the original Pareto-optimal sets for MI and SSVEP tasks)
Output:
PS_MI and PS_SSVEP (the updated Pareto-optimal sets for MI and SSVEP tasks)

Generate parent population (parent_POP) by tournament selection operator from POP;
$i = 1$, offspring_POP $= \emptyset$;
while $i <$size(parent_POP) **do**
  Select parent_POP(i) and parent_POP(i+1) as parent1 and parent2, respectively;
  if task_label(parent1)=task_label(parent2) **then**
    Generate two offspring individuals (offspring1 and offpring2) according to partial-mapped crossover operator;
    task_label(offspring1)=task_label (parent1), task_label(offspring2) = task_label (parent2);
  else
    Obtain r_1 which is randomly generated from [0, 1];
    if r_1 is smaller than the predefined crossover probability **then**
      Generate two offspring individuals (offspring1 and offpring2) according to partial-mapped crossover operator;
    else
      Obtain r_2 which is randomly generated from [0, 1];
      if r_2 is smaller than the predefined mutation probability **then**
        Generate two offspring individuals (offspring1 and offpring2) according to flip-bit mutation operator;
      end if
    end if
  end if
  task_label(offspring1) = task_label (parent1), task_label(offspring2) = task_label (parent2);
  offspring_POP = offspring_POP ∪ { offspring1,offspring2 };
end while
temp_POP = POP ∪ offspring_POP;
Obtain the next POP by selecting the fittest individuals from temp_POP
Update PS_MI as the non-dominated solutions from the original PS_MI and the individuals with task_label = 1 in POP and update PS_SSVEP as the non-dominated solutions from the original PS_SSVEP and the individuals with task_label = 2 in POP.

set outputted by the whole algorithm consists of a set of non-dominated solutions which aim to make a compromise among *MAR*, *SAR*, and *NC*. To improve the searching efficiency, a local searching operator, which is based on a decision variable analysis strategy, is introduced in this stage.

TABLE 1 | Parameter settings of algorithms.

Algorithm	Parameter settings
EMMOA	Population size: 100; crossover parameter: 0.8; mutation probability: 0.2
NSGA-II	Population size: 100
MOEA/D	Population size: 105
MOPSO	Population size: 100

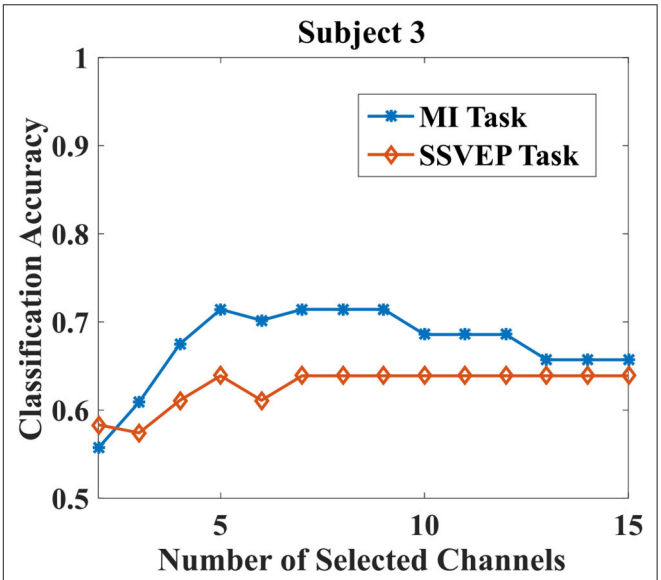


FIGURE 3 | Classification accuracies for MI and SSVEP tasks with different numbers of selected channels.

The main idea of the proposed decision variable analysis operator is to divide the decision variables into three groups: add_group, delete_group, and invalid_group. The division of the first 11 and the last 4 decision variables are based on PS_SSVEP and PS_MI, respectively. Take the jth ($1 \leq j \leq 11$) decision variable for an example. At first, initialize *flag* to 0. For a non-dominated solution (x_i) in PS_SSVEP, if the classification accuracy is raised with $x_i(j) = 0$, then one can consider the performance of x_i will be improved without selecting the jth channel. In this case, $flag = flag - 1$. If $x_i(j) = 1$ gets a better classification accuracy, then $flag = flag + 1$. If the classification accuracy is not affected by $x_i(j)$, then the value of *flag* will remain unchanged. All the members in PS_SSVEP will be executed to get the final value of *flag*.The jth decision variable will be divided into different groups according to the value of *flag*. The detailed procedure of the proposed decision variable analysis operator is given in **Algorithm 2**.

After decision variable analysis, the local searching operator is implemented on PS_MI and PS_SSVEP. The detailed procedure of the proposed local searching operator is shown in **Algorithm 3**. For the Pareto-optimal solutions in PS_MI, the searching directions are obtained according to the types

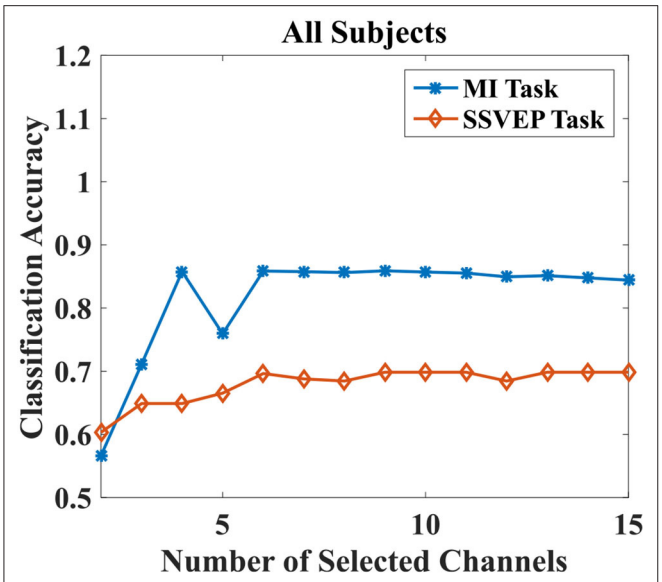


FIGURE 4 | Average classification accuracies for MI and SSVEP tasks with different numbers of selected channels for all subjects.

of the last four decision variables (lines 2–11 in **Algorithm 3**). For the Pareto-optimal solutions in PS_SSVEP, the searching directions are obtained according to the types of the first eleven decision variables (lines 12–21 in **Algorithm 3**). The current Pareto-optimal set of EMMOA will be acquired by selecting the non-dominated solutions in the original Pareto-optimal set and newPOP obtained by the local searching operator.

Algorithm 2 Detailed procedure of division variable analysis operator.

Input:
PS_MI and PS_SSVEP
n1(number of the solutions in PS_MI), n2(number of the solutions in PS_SSVEP)
Output:
add_group, delete_group, and invalid_group

add_group=∅, delete_group = ∅, invalid_group = ∅;
for $j = 1$ TO 11 **do**
  $flag = 0$;
  for $i = 1$ TO n1 **do**
    x is the ith solution in PS_MI, $x1 = x$;
    x_MAR is the MI classification accuracy of the ith solution in PS_MI;
    if $x1(j) = 0$ **then**
      $x1(j) = 1$ and $x1_MAR$ is the MI classification accuracy of $x1$;
      if $x_MAR < x1_MAR$ **then**
        $flag = flag + 1$;
      else if $x_MAR > x1_MAR$ **then**
        $flag = flag - 1$;
      end if
    else
      $x1(j) = 0$ and $X1_MAR$ is the MI classification accuracy of $x1$;
      if $x_MAR < x1_MAR$ **then**
        $flag = flag - 1$;
      else if $x_MAR > x1_MAR$ **then**
        $flag = flag + 1$;
      end if
    end if
  end for
  if $flag > 0$ **then**
    add_group = add_group ∪ {i};
  else if flag< 0 **then**
    delete_group = delete_group ∪ {i};
  else
    invalid_group = invalid_group ∪ {i};
  end if
end for

3. EXPERIMENT AND RESULT ANALYSIS

3.1. Experimental Setup

This section contained two experiments. The first one aimed to demonstrate whether the proposed channel selection algorithm (EMMOA) could obtain a smaller set of channels with as little sacrifice as possible in MI and SSVEP classification accuracy. The second one aimed to evaluate the effectiveness of EMMOA by comparing with three widely used multiobjective optimization algorithms, including NSGA-II (Deb et al., 2002), MOEA/D (Zhang and Li, 2007), and MOPSO (Coello et al., 2004). The number of total function evaluations was 10000 in a single run for all 4 algorithms. The detailed parameter settings of algorithms were given in **Table 1**.

3.2. Results and Analysis

The purpose of the proposed EMMOA is to find the optimum number of channels that would result in a promising performance for both MI and SSVEP tasks. Take subject 3 as an example, **Figure 3** illustrated the classification accuracies for MI and SSVEP tasks with different numbers of selected channels. It can be observed from **Figure 3**, the classification accuracy for MI task tends to become higher at first and then become lower as the number of selected channels increases. The classification accuracy for SSVEP task tends to be stable when the number of selected channels. **Figure 4** illustrated the average classification accuracies for all 7 subjects with different numbers of selected channels. **Figure 4** demonstrated that the larger number of selected channels cannot lead to better performance for MI and SSVEP tasks. This may indicate the necessity of selecting a promising subset of channels.

Table 2 demonstrated the classification accuracy for MI and SSVEP tasks by using all channels and the reduced set of EEG channels selected by the proposed EMMOA. As **Table 2** shows, EMMOA can find a smaller set of channels without depredating the classification performance for all 7 subjects. For example, the classification accuracies for MI and SSVEP tasks using 5 channels

Algorithm 3 Detailed procedure of local searching operator.

Input:
add_group, delete_group, and invalid_group
n1 (number of the solutions in PS_MI), n2 (number of the solutions in PS_SSVEP)
Output:
newPOP

newPOP=∅;
for $i = 1$ TO n1 **do**
  x is the ith solution in PS_MI, $x1 = x$;
  r is a random integer from [12, 15];
  if $r \in$ add_group and $x(i) = 0$ **then**
    $x1(i) = 1$;
  else if $r \in$ delete_group and $x(i) = 1$ **then**
    $x1(i) = 0$;
  end if
  newPOP = newPOP $\cup x1$;
end for
for $i = 1$ TO n2 **do**
  x is the ith solution in PS_SSVEP, $x1 = x$;
  r is a random integer from [1, 11];
  if $r \in$ add_group and $x(i) = 0$ **then**
    $x1(i) = 1$;
  else if $r \in$ delete_group and $x(i) = 1$ **then**
    $x1(i) = 0$;
  end if
  newPOP = newPOP $\cup x1$;
end for

TABLE 2 | Classification accuracies for each subject by using all channels and the selected channels obtained by EMMOA.

Subject	All Channels		Using EMMOA		
	MAR (%)	*SAR* (%)	*MAR* (%)	*SAR* (%)	Number of selected channels
Subject-1	1	0.6111	1	0.6667	4
Subject-2	0.5824	0.5	0.6574	0.5833	5
Subject-3	0.6571	0.6388	0.7143	0.6389	5
Subject-4	1	0.8611	1	0.9444	4
Subject-5	0.67	0.5833	0.75	0.5833	4
Subject-6	1	0.7222	1	0.7222	5
Subject-7	1	0.9722	1	1	4

selected by EMMOA are better than those using all channels. For subject 6, the classification performance using all channels is same to that of using 5 channels. However, a smaller number of channels is certainly more convenient for real-life applications. **Table 3** gives the average classification accuracies for all subjects by using all channels and the Pareto-optimal set obtained by EMMOA in a single run. It can be observed from **Table 3**, the classification performance by using 6, 7, and 8 channels selected by EMMOA is better than using all channels. Therefore, the experimental results demonstrated the effectiveness of using the proposed channel selection algorithm.

For further comparison of the proposed EMMOA with other classical multi-objective evolutionary algorithms, three algorithms, including NSGA-II, MOEA/D and MOPSO, are adopted in this section. All the comparative algorithms have been running 30 times to get the statistical results, which are

TABLE 3 | Average classification accuracies for all subjects by using all channels and the pareto-optimal set obtained by EMMOA for all subjects.

All channels		PS by EMMOA		
MAR (%)	*SAR* (%)	*MAR* (%)	*SAR* (%)	Number of selected channels
0.8442	0.6984	0.5661	0.6032	2
		0.5661	0.6944	3
		0.8490	0.6032	3
		0.8561	0.5893	3
		0.8561	0.6944	4
		0.8597	0.6012	4
		0.5661	0.6984	5
		0.8597	0.6944	5
		0.8561	0.6984	6
		0.8581	0.6984	7
		0.8611	0.6468	7
		0.8611	0.6984	8

TABLE 4 | Average results obtained by all algorithms.

Algorithm	Average *MAR* (%)	Average *SAR* (%)	Average number of selected channels	Average rank
NSGA-II	0.7606 (4)	0.6601 (2)	4.48 (1)	2.33
MOEA/D	0.8480 (1)	0.6586 (2)	6.67 (4)	2.67
MOPSO	0.7684 (3)	0.6553 (4)	4.83 (3)	3.33
EMMOA	0.7969 (2)	0.6653 (1)	4.57 (2)	1.67

TABLE 5 | Maximum results obtained by all algorithms.

Algorithm	Maximum *MAR* (%)	Maximum *SAR* (%)	Number of selected channels
NSGA-II	**0.8561**	0.6032	3
MOEA/D	0.8362	0.6032	
MOPSO	0.5661	**0.6944**	
EMMOA	**0.8561**	**0.6944**	
NSGA-II	0.8597	0.6944	5
MOEA/D	0.8526	0.6548	
MOPSO	0.8597	**0.6984**	
EMMOA	**0.8611**	**0.6984**	
NSGA-II	**0.8611**	**0.6984**	7
MOEA/D	0.8397	**0.6984**	
MOPSO	0.8597	**0.6984**	
EMMOA	**0.8611**	**0.6984**	

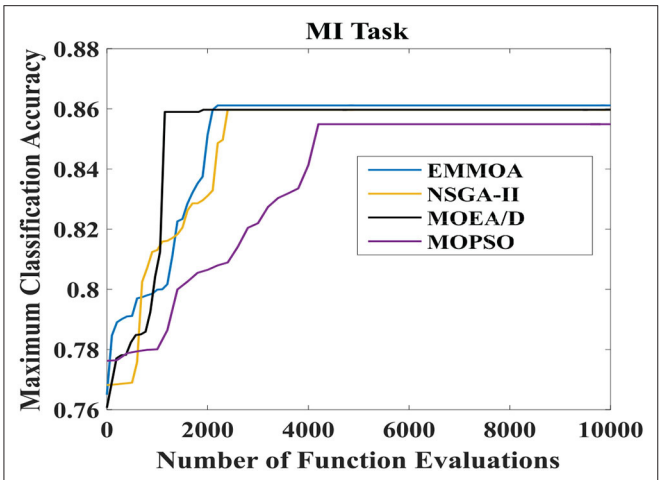


FIGURE 5 | Convergence of four algorithms for MI task.

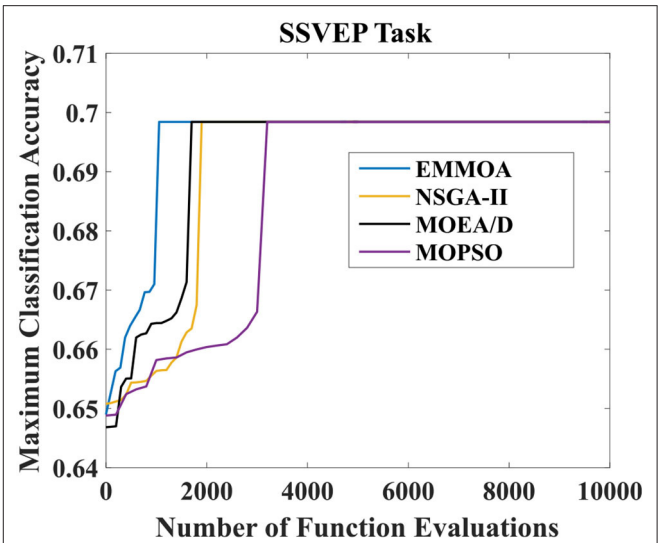


FIGURE 6 | Convergence of four algorithms for SSVEP task.

shown in **Table 4**. In **Table 4**, the numbers in brackets indicate the rank of the corresponding algorithm in terms of average MI accuracy, average SSVEP accuracy, and average number of selected channels, respectively. A smaller rank value of an algorithm means a better performance achieved by the algorithm. In the last column of **Table 4**, EMMOA obtained the smallest (best) average rank. In other words, the proposed channel selection algorithm achieved the best performance in terms of average rank. **Table 5** gives the maximum average classification accuracies for MI and SSVEP tasks in the PS obtained by all the 4 algorithms with different numbers of selected channels. The numbers in boldface give the best accuracy rates in different cases. It can be observed from **Table 5**, EMMOA obtained the best classification accuracies for all 3 cases.

The results in **Tables 4**, **5** demonstrated the effectiveness of the proposed algorithm. This may be because the multitasking mechanism adopted in the first stage of EMMOA help the algorithm find more promising PS for each task efficiently. On the other hand, with the help of decision variables analysis, the local searching strategy in the second stage makes a better compromise among MI classification accuracy, SSVEP classification accuracy, and the number of selected channels.

Figures 5, **6** illustrated the convergence of four algorithms for MI and SSVEP tasks, respectively. In **Figures 5**, **6**, the x-axis represented the number of function evaluations and the y-axis displayed the maximum classification accuracy averaged over all the 7 subjects. As **Figure 5** showed, the proposed EMMOA obtained the second-best and the best convergence speed for MI and SSVEP tasks among four algorithms, respectively. This is because the decision variable analysis strategy adopted in EMMOA helped the algorithm searching more efficiently.

4. DISCUSSION

It can be observed from **Figure 4**, the best classification accuracy can be achieved when the number of selected channels is not very large. This observation can be validated by the results demonstrated in **Table 3**. The classification accuracies for both MI and SSVEP tasks are higher than 0.65 when the number of selected channels is larger than 3. As **Figure 4** showed, the improvement in classification accuracy is not obvious by adopting 4 channels when compared with using more than 4 channels. Especially for MI task, the classification accuracy by using 4 channels is even better than using all 15 channels. The phenomenon is in agreement with the conclusion in Moubayed et al. (2010), which showed that some channels may bring artifacts and then lead to a degeneration in classification accuracy.

TABLE 6 | Results outputted by decision variable analysis operator.

First 11 channels (for MI Task)		Last 4 channels (for SSVEP Task)	
Middle Stage of EMMOA			
add_group	{2, 3, 5, 8, 9}	add_group	{12, 13, 14, 15}
delete_group	{1, 7, 10, 11}	delete_group	{}
invalid_group	{6}	delete_group	{}
Last stage of EMMOA			
add_group	{2, 3, 4, 7, 8, 9}	add_group	{12, 13, 14, 15}
delete_group	{1, 5, 6, 10, 11}	delete_group	{}
invalid_group	{}	delete_group	{}

In the proposed algorithm, the decision variable analysis operator played a vital role in the second stage of EMMOA, since it provided searching directions for the next local searching operator. **Table 6** showed the results outputted by the decision variable analysis operator in the middle and last stages of the optimization process by using EMMOA. add_group, delete_group and invalid_group contains a set of channels that are considered to be profitable, disadvantageous, and invalid for the corresponding classification task. As **Table 6** shows, most results obtained by the decision variable analysis operator in

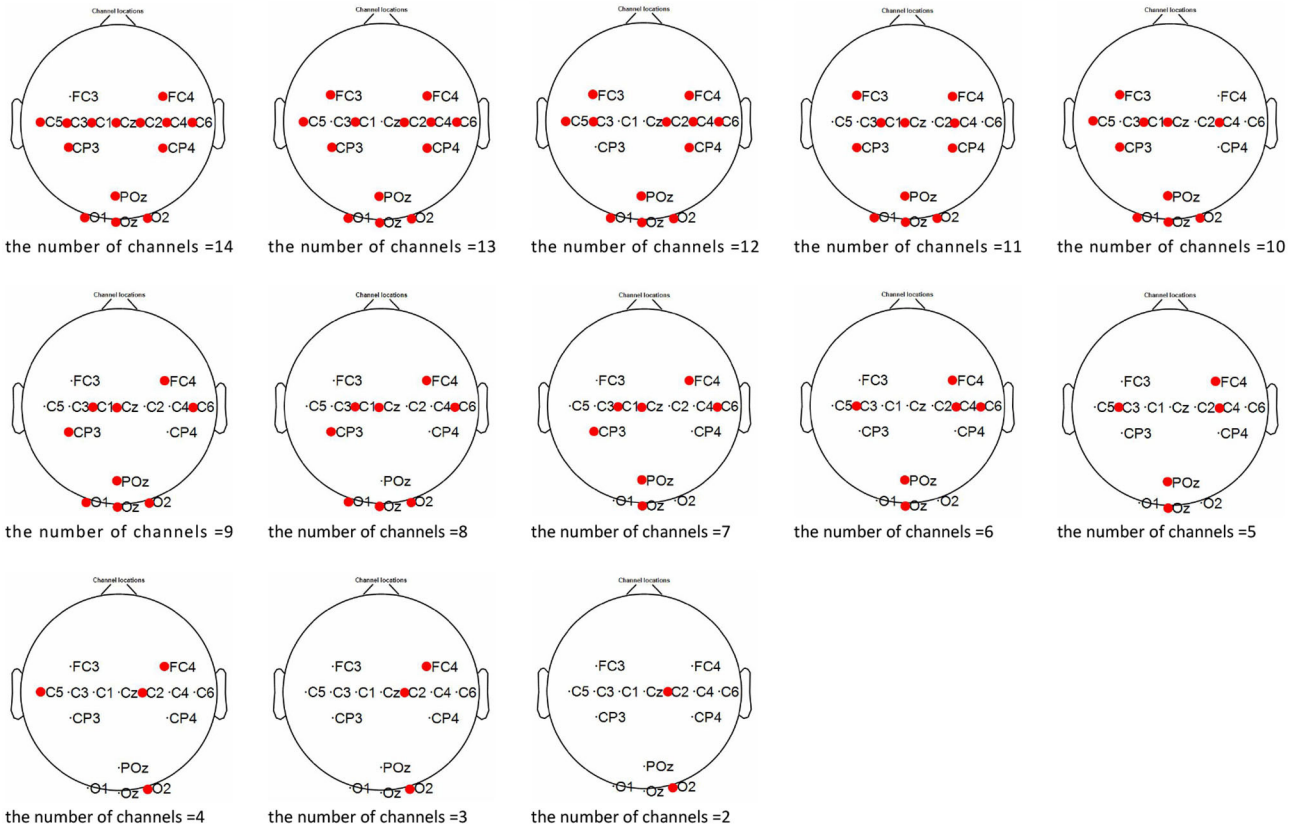


FIGURE 7 | The distribution diagrams in conditions of different numbers of selected channels.

the middle stage overlap with those obtained in the last stage. However, the results in different stages are not exactly the same. This phenomenon may indicate the uncertainty of the decision variable analysis operator to some extent. As described in section 2.4, the performance of the second stage of EMMOA depends on the results of the proposed decision variable analysis operator. Therefore, the second stage may not obtain a promising output if the decision variable analysis operator cannot get an appropriate division of the decision variables, especially in the early stage of the evolution process. How to improve the stability of the proposed decision variable analysis operator will be the next work of this paper.

As shown in **Figure 7**, the distribution diagrams were illustrated in conditions of different numbers of selected channels. Evidently, the electrodes were first eliminated from motor cortex when the number was sufficient for MI and SSVEP recognition. And the quantity of channels were close between central area and occipital area along with the decline of total number. Meanwhile, the priority of selecting was unclear in these two regions because of the alternate quantitative superiority. However, the feature information needed to be reserved those extracted from left- and right- motor cortex for spatial filtering. Thus, the limitation of channel number in the occipital area might be firstly considered as the kernel factor of EMMOA.

5. CONCLUSION

In our study, a multitasking-based multiobjective evolutionary algorithm (EMMOA) was proposed to select appropriate channels for these two classification tasks at the same time. Moreover, a two-stage framework was introduced to balance the number of selected channels and the classification accuracy in the proposed algorithm. The experimental results verified the feasibility of multiobjective optimization methodology for channel selection of hybrid BCI tasks.

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 Ethics Committee of Tongji University. The patients/participants provided

their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TL designed the methodology of data process and performed the data analysis. TL and ZX organized the data and wrote the manuscript. LC and GT reviewed and edited the manuscript. All authors read and approved the submitted manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant nos. 61806122 and 62102242).

REFERENCES

Alcaide-Aguirre, R. E., and Huggins, J. E. (2014). Novel hold-release functionality in a p300 brain-computer interface. *J. Neural Eng.* 11:066010. doi: 10.1088/1741-2560/11/6/066010

Bai, L., Lin, W., Gupta, A., and Song, Y. S. (2021). From multitask gradient descent to gradient-free 15 evolutionary multitasking: a proof of faster convergence. *IEEE Trans. Cybern*. doi: 10.1109/TCYB.2021.3052509. [Epub ahead of print].

Chicano, F., Sutton, A. M., Whitley, L. D., and Alba, E. (2015). Fitness probability distribution of bit-flip mutation. *Evol. Comput*. 23, 217–248. doi: 10.1162/EVCO_a_00130

Coello, C. A. C., Pulido, G. T., and Lechuga, M. S. (2004). Handling multiple objectives with particle swarm optimization. *IEEE Trans. Evol. Comput*. 8, 256–279. doi: 10.1109/TEVC.2004.826067

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Trans. Evolut. Comput*. 6, 182–197. doi: 10.1109/4235.996017

Gong, M., Tang, Z., Li, H., and Zhang, J. (2019). Evolutionary multitasking with dynamic resource allocating strategy. *IEEE Trans. Evolut. Comput*. 23, 858–869. doi: 10.1109/TEVC.2019.2893614

Gupta, A., Ong, Y. S., and Feng, L. (2015). Multifactorial evolution: toward evolutionary multitasking. *IEEE Trans. Evolut. Comput*. 20, 343–357. doi: 10.1109/TEVC.2015.2458037

Gupta, A., Ong, Y. S., Feng, L., and Tan, K. C. (2016). Multiobjective multifactorial optimization in evolutionary multitasking. *IEEE Trans. Cybern*. 47, 1652–1665. doi: 10.1109/TCYB.2016.2554622

Hasan, B. A. S., Gan, J. Q., and Zhang, Q. (2010). "Multi-objective evolutionary methods for channel selection in brain-computer interfaces: some preliminary experimental results," in *IEEE Congress on Evolutionary Computation* (Barcelona: IEEE), 1–6.

Ismkhan, H., and Zamanifar, K. (2015). Study of some recent crossovers effects on speed and accuracy of genetic algorithm, using symmetric travelling salesman problem. *arXiv preprint arXiv*: 1504.02590.

Jin, J., Liu, C., Daly, I., Miao, Y., Li, S., Wang, X., et al. (2020). A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2153–2163. doi: 10.1109/TNSRE.2020.3020975

Jin, J., Miao, Y. Y., Daly, I., Zuo, C. L., Hu, D. W., and Cichocki, A. (2019). Correlation-based channel selection and regularized feature optimization for mi-based bci. *Neural Netw*. 118, 262–270. doi: 10.1016/j.neunet.2019.07.008

Kee, C. Y., Ponnambalam, S. G., and Loo, C. K. (2015). Multi-objective genetic algorithm as channel selection method for p300 and motor imagery data set. *Neurocomputing* 161, 120–131. doi: 10.1016/j.neucom.2015.02.057

Kevric, J., and Subasi, A. (2017). Comparison of signal decomposition methods in classification of eeg signals for motor-imagery bci system. *Biomed. Signal Process. Control* 31, 398–406. doi: 10.1016/j.bspc.2016.09.007

Ko, L. W., Lin, S. C., Song, M. S., and Komarov, O. (2014). "Developing a few-channel hybrid bci system by using motor imagery with ssvep assist," in *2014 International Joint Conference on Neural Networks (IJCNN)* (Beijing: IEEE), 4114–4120.

Lin, Z., Zhang, C., Wu, W., and Gao, X. (2006). Frequency recognition based on canonical correlation analysis for ssvep-based bcis. *IEEE Trans. Biomed. Eng.* 53, 261–274. doi: 10.1109/TBME.2006.886577

Long, J., Li, Y., Wang, H., Yu, T., Pan, J., and Li, F. (2012). A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 720–729. doi: 10.1109/TNSRE.2012.2197221

Moubayed, N. A., Hasan, B. A. S., Gan, J. Q., and Petrovski, A. (2010). "Binary-sdmopso and its application in channel selection for brain-computer interfaces," in *2010 UK Workshop on Computational Intelligence (UKCI)* (Colchester: IEEE), 1–6.

Muller-Gerking, J., Pfurtscheller, G., and Flyvbjerg, H. (2004). Designing optimal spatial filters for single-trial eeg classification in a movement task. *Clin. Neurophysiol*. 110, 787–798. doi: 10.1016/S1388-2457(98)00038-8

Pfurtscheller, G., and da Silva, F. H. L. (1999). *Handbook of Electroencephalography and Clinical Neurophysiology-Event-Related Desynchronization*. Amsterdam: Elsevier.

Qiu, Z., Jin, J., Lam, H. K., Zhang, Y., and Cichocki, A. (2016). Improved sffs method for channel selection in motor imagery based bci. *Neurocomputing* 207, 519–527. doi: 10.1016/j.neucom.2016.05.035

Ravi, A., Pearce, S., Zhang, X., and Jiang, N. (2019). "User-specific channel selection method to improve ssvep bci decoding robustness against variable inter-stimulus distance," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Francisco, CA: IEEE), 283–286.

Su, J. R., Wang, J. G., Xie, Z. T., Yao, Y., and Liu, J. (2019). "A method for eeg contributory channel selection based on deep belief network," in *IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)* (Dali: IEEE), 1247–1252.

Sun, H., Jin, J., Kong, W. Z., Zuo, C. L., Li, S. R., and Wang, X. Y. (2021). Novel channel selection method based on position priori weighted permutation entropy and binary gravity search algorithm. *Cogn. Neurodyn*. 15, 141–156. doi: 10.1007/s11571-020-09608-3

Wang, B. C., Li, H. X., Zhang, Q., and Wang, Y. (2018). Decomposition-based multiobjective optimization for constrained evolutionary optimization. *IEEE Trans. Syst. Man Cybern. Syst*. 51, 574–587. doi: 10.1109/TSMC.2018.2876335

Yang, Y., Bloch, I., Chevallier, S., and Wiart, J. (2016). Subject-specific channel selection using time information for motor imagery brain-computer interfaces. *Cogn. Comput*. 8, 505–518. doi: 10.1007/s12559-015-9379-z

Zhang, J., Yan, C., Cao, L., and Gong, X. (2017). "Optimal channel set selection for ssvep-based bci using spatial temporal correlation," in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (Guilin: IEEE), 2038–2042.

Zhang, Q., and Li, H. (2007). Moea/d: a multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. Evolut. Comput*. 11, 712–731. doi: 10.1109/TEVC.2007.892759

Zhang, Y., Zhen, Y., and Lu, T. T. (2016). "A fuzzy c-means clustering based tournament selection for multiobjective optimization," in *2016 IEEE Congress on Evolutionary Computation (CEC)* (Vancouver, BC: IEEE), 2446–2453.

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

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

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

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 15 October 2021
doi: 10.3389/fnins.2021.760987

Automatic Seizure Classification Based on Domain-Invariant Deep Representation of EEG

Xincheng Cao[1,2], Bin Yao[1,2], Binqiang Chen[1,2], Weifang Sun[3]* and Guowei Tan[4]**

[1] School of Aerospace Engineering, Xiamen University, Xiamen, China, [2] Shenzhen Research Institute of Xiamen University, Shenzhen, China, [3] College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China, [4] Xiamen Key Laboratory of Brain Center, Department of Neurosurgery, The First Affiliated Hospital of Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Shaohui Zhang,
Dongguan University of Technology, China
Xingwu Zhang,
Xi'an Jiaotong University, China
Junjie Ye,
Xidian University, China

***Correspondence:**
Bin Yao
yaobin@xmu.edu.cn
Weifang Sun
swf@wzu.edu.cn
Guowei Tan
gwtanxm@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *19 August 2021*
Accepted: *13 September 2021*
Published: *15 October 2021*

Citation:
Cao X, Yao B, Chen B, Sun W and Tan G (2021) Automatic Seizure Classification Based on Domain-Invariant Deep Representation of EEG. Front. Neurosci. 15:760987. doi: 10.3389/fnins.2021.760987

Accurate identification of the type of seizure is very important for the treatment plan and drug prescription of epileptic patients. Artificial intelligence has shown considerable potential in the fields of automated EEG analysis and seizure classification. However, the highly personalized representation of epileptic seizures in EEG has led to many research results that are not satisfactory in clinical applications. In order to improve the clinical adaptability of the algorithm, this paper proposes an adversarial learning-driven domain-invariant deep feature representation method, which enables the hybrid deep networks (HDN) to reliably identify seizure types. In the train phase, we first use the labeled multi-lead EEG short samples to train squeeze-and-excitation networks (SENet) to extract short-term features, and then use the compressed samples to train the long short-term memory networks (LSTM) to extract long-time features and construct a classifier. In the inference phase, we first adjust the feature mapping of LSTM through the adversarial learning between LSTM and clustering subnet so that the EEG of the target patient and the EEG in the database obey the same distribution in the deep feature space. Finally, we use the adjusted classifier to identify the type of seizure. Experiments were carried out based on the TUH EEG Seizure Corpus and CHB-MIT seizure database. The experimental results show that the proposed domain adaptive deep feature representation improves the classification accuracy of the hybrid deep model in the target set by 5%. It is of great significance for the clinical application of EEG automatic analysis equipment.

Keywords: electroencephalography, seizure classification, deep learning, domain-invariant representation, hybrid deep model

INTRODUCTION

According to the World Health Organization (WHO), nearly 50 million people suffer from epilepsy worldwide, and it is estimated that 2.4 million people are diagnosed with epilepsy annually (World Health Organization, 2019). In order to start anti-epileptic drugs or treatment, timely and accurate diagnosis of epilepsy is very important. Nowadays, electroencephalogram (EEG) plays a significant role in epilepsy diagnosis. EEG is the record of electrical

activity collected through electrodes on the scalp to provide spatial and temporal information of the brain. The neurologist checks the EEG for abnormal brain electrical activity, identify the type of seizure, and perform pathological analysis. However, manual analysis of EEG records is very time-consuming, especially with the increase of outpatient routine EEG and inpatient long term monitoring, and neurologists often spend several hours a day to review EEG (Krumholz et al., 2007). Therefore, the intelligent EEG analysis and classification algorithm has become the key to improving the level of epilepsy treatment.

The automatic analysis process of EEG can be roughly divided into several steps such as signal acquisition, preprocessing, feature extraction, and pattern recognition. Due to random and complex neural activity, EEG signals are inherently non-linear, non-stationary, and highly random. The filtering and noise reduction of the original EEG signal is very important for artifact recognition and feature extraction (Chen et al., 2019). According to clinical experience, epilepsy can cause changes in the waveform and frequency of EEG. Therefore, time-frequency analysis tools such as wavelet transform are widely used in EEG preprocessing and feature engineering. Mahjoub adapted tunable-Q wavelet transform and multivariate empirical mode decomposition to the preprocessing and feature extraction of EEG (Mahjoub et al., 2020). Slimen removed artifact by Savitzky–Golay filter and multi-scale principal component analysis (Slimen et al., 2020). He proposed a periodic overlapping group sparsity method for sparse feature extraction (He et al., 2018). Hussain proposed a novel entropy index permutation fuzzy entropy that significantly improves the accuracy of a variety of machine learning algorithms (Hussain et al., 2019). Gabara uses the CHB-MIT database to study the machine learning algorithm for identifying EEG during seizures. The experimental results suggest that indicators such as Fractal Dimension, Fluctuation Index, Variation Coefficient, and Kurtosis are more effective (Gabara et al., 2020).

Feature engineering reveals the difference between epileptic EEG and normal EEG, and also reduces the dimensionality of the data, so that machine learning models can be used to classify samples. Since the database disclosed in the early years only had two types of labels, normal and seizure, support vector machines (SVM), an algorithm particularly suitable for two classification problems, have been widely used. RM utilized SVM to identify seizure EEG signal (Aileni et al., 2019). Li proposed a novel EEG feature matrix generation algorithm based on fast Fourier transform, then used principal component analysis network for dimensionality reduction, and finally used SVM for identification, which were implemented on the Bonn database and the CHB-MIT database. The accuracy of 99 and 98% are achieved (Li and Chen, 2021). He proposed a diagnosis method based on sparse demodulation operator (He et al., 2019). Subasi used particle swarm optimization to determine the optimum parameters of SVM and increased the accuracy of epileptic seizure to 99.38% (Subasi et al., 2019). Experiments conducted by Shabarinath show that the combination of feature extraction *via* discrete wavelet transform and seizure identification *via* extreme learning machine achieved the highest accuracy, reported to be 90.1% (Shabarinath et al., 2019).

Despite many research cased combing feature engineering and machine learning have achieved high recognition accuracy on public databases. The modeling capabilities of machine learning models are limited, and severe degradation has occurred on large databases (Roy et al., 2019). The exciting thing is that, owing to the accumulation of database, deep learning models that can process big data through self-learning have been proven to be similar to humans in biomedical data analysis (Rajpurkar et al., 2017; Yıldırım et al., 2018; Arsenovic et al., 2020). Ozal uses a one-dimensional convolutional neural network (CNN) to analyze single-channel digital records to identify abnormal EEG, and achieves an accuracy of 79% on the Temple University Hospital EEG Abnormal Corpus (Yildirim et al., 2020). Liu applied smoothing and collar technique to the outputs of CNN and increased the sensitivity of seizure detection to 97% (Liu et al., 2019). Wang calculated the sub-band mean amplitude spectrum map (MAS) of the multi-channel EEG as a two-dimensional representation of the EEG and then trained the CNN to recognize the types of EEG samples (preictal, ictal, interictal), and the overall accuracy reached 92.77% (Wang Y. et al., 2019). Yang uses a convolutional autoencoder to extract features from EEG records and then uses a long short-term memory network (LSTM) to detect seizure (Li et al., 2020b). Abdelhameed uses multi-lead EEG as a monitoring method, uses a two-dimensional convolutional autoencoder to extract features from a time-channel two-dimensional matrix, and then uses LSTM to identify seizures. The accuracy achieved on the CHB-MIT database is reported above 98 (Abdelhameed and Bayoumi, 2021). Li uses the squeeze excitation module that can model the relationship between channels to adaptively select the EEG leads processed by the deep model. For the Temple University Hospital EEG Seizure Corpus (TUSZ), the accuracy of the proposed PSNE to identify epilepsy types is reported to reach 92% (Li et al., 2020a).

The series of deep models mentioned above achieve high accuracy on the database used. However, the manifestation of epilepsy in EEG is affected by various physiological indicators, and it is difficult to collect all possible databases. When the object monitored by the EEG classification model differs from the training data, the classification accuracy will inevitably decrease. The exciting thing is that deep learning experts provide a possible solution, which is transfer learning (Wang S.-H. et al., 2019; Cao et al., 2021; Zhuang et al., 2021). There have been studies exploring the application of transfer learning to ECG classification (Alghamdi et al., 2020; Naz et al., 2021). Wang proposed a continuous index adaptive algorithm, which extended the domain adaptive learning to the continuously changing domain (Wang et al., 2020). Owing to the powerful feature extraction of the deep model, deep domain adaptation has succeeded in machine vision, fault diagnosis, and other fields. Constraints were imposed on the deep representation, such as the maximum mean discrepancy (Cao et al., 2020), to align the distribution of the target domain samples and the source domain samples in the deep feature space, thereby improving the classification accuracy of the classifier in the target domain. Yin applies domain adaptation to personalized ECG monitoring combined with IR-UWB radar (Yin et al., 2019). Wang proposed

an arrhythmia heartbeat classification method for ECG based on unsupervised domain adaptation.

In order to improve the reliability of EEG monitoring equipment in detecting seizure and classification in clinical applications, a hybrid deep network is proposed in this work, and a domain adaptive training method is proposed accordingly. First, we use the labeled samples in the database to train hybrid deep networks (HDN) to construct the deep feature map of EEG. Then, we use the unlabeled samples of the target patient to carry out domain adaptive learning, and fine-tune the HDN to improve the domain invariance of the deep re-representation. We realize the accurate classification of seizure.

METHODS

Overview of the Proposed Method and Hybrid Deep Networks

The pipeline of the proposed method is illustrated in **Figure 1**, and the utilized HDN consists of three sub-networks, 1-D squeeze-and-excitation networks (SENet), LSTM, and Multilayer Perceptron (MLP). SENet and LSTM are responsible for extracting deep features of multiple time spans from EEG, and MLP is responsible for identifying seizure types based on the deep features of EEG. Firstly, the EEG monitoring device is used to collect the EEG of patients and then store it as a channel-time two-dimensional matrix. Taking the current moment as the origin, eight short-term EEG samples with a duration of 2 s are successively cropped to the previous. Firstly, we input the 2-s multi-lead EEG matrix into the 1-D CNN to extract deep features.

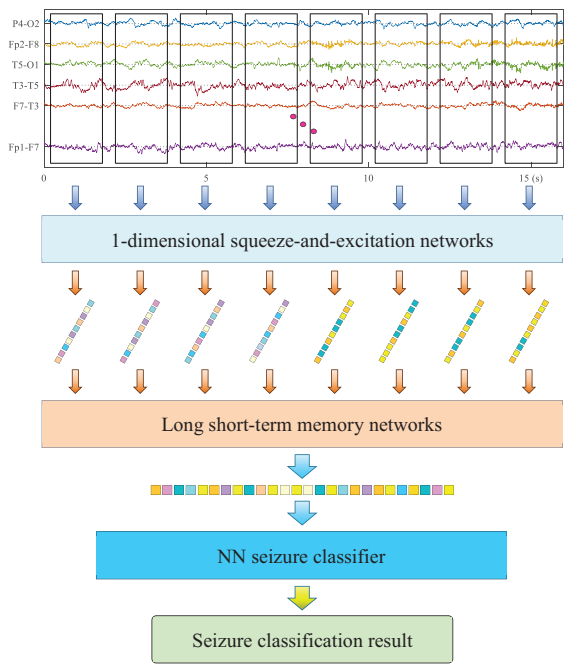


FIGURE 1 | The architecture of the utilized hybrid deep networks (HDN).

The utilized 1-D CNN has as many input channels as EEG leads, and each input channel accepts one lead of EEG data. Then, the short-term deep feature vectors of the eight short-term EEG samples are sequentially input to the LSTM to extract long-term deep features. Finally, a classifier is used to identify the health status of EEG and predict the type of seizure.

Short-Term Feature Extraction *via* Squeeze-and-Excitation Networks

According to the professional knowledge of neurology, epilepsy is manifested as changes in waveform and frequency in EEG. The superior performance of CNN in extracting waveform and frequency features has been widely verified in the field of machine vision and natural language processing. On the other hand, different types of seizure will cause changes in different leads, for example, local seizure only shows variation in several leads. Therefore, the deep network used should be able to model the relationship between channels and identify the differences between different leads. For this reason, this research adapted SENet to extract deep features from short-term EEG samples and reduce the dimensionality. The utilized SENet consists of five SE-Res blocks connected in series. The number of channels of each block is 20, 40, 80, 160, and 160. The length of the convolution kernel of each block is 16, 8, 5, 3, and 3.

The utilized SE-Res block constants consist of two convolutional layer, a global maximum pooling layer, an embedded fully connected neural network (NN), a scale block, a skip path, and an Add operation. The convolutional layer extracts feature from each channel and adjusts the number of channels. The global maximum pooling maps the feature maps to channel indices. The embedded NN takes the output of the global pooling layer as the input layer and then outputs the activation value of each channel. In order to simplify training and reduce calculations, the embedded NN uses only one hidden layer. The inventor of the SE module proposes to set the number of neurons in the hidden layer to one-sixteenth of the number of input channels. Considering the small number of channels in each module in this study, the number of neurons in the hidden layer is consistently set to 20. The first fully connected layer uses the *ReLU* activation function. The second fully connected layer uses the sigmoid activation function to map the activation value of the channel to the interval [0,1]. The scaling block assigns the activation value of the channel to each feature map. The structure of all SE-Res blocks in the model remains the same.

Long-Term Feature Extraction *via* Long Short-Term Memory Network

The representation of seizure in EEG has strong individuality. It is difficult to balance the accuracy and generalization ability with a single short-term EEG sample to identify the seizure type of the patient. To this end, we use LSTM to extract the long-term features of EEG after SENet. Three LSTM modules are stacked to form a deep network, and after each LSTM module, the pooling layer is used for down-sampling. Inside the LSTM, a four-layer fully connected neural network is used to control each gate.

Suppose that x^{t-1} and x^t are the feature vectors extracted by SENet from the short-term EEG samples at time $t-1$ and time t, respectively. h^{t-1} is the output of LSTM with the input of x^{t-1}, which is also called the hidden cell state. Three neural networks are used determine which historical information should be inherited into the processing of the feature vector at the current moment, which historical information should be updated, and extract features from the current input feature vector. The forget gate and input gate in LSTM are merged into an update gate u^t, which determines which parts of the current information should be output and at the same time determines which historical information should be forgotten $1-u^t$. One sub-network is reduced, and calculation is easier. The output of LSTM and the operation of each gate in it are defined by the following equations:

$$h^t = (1-u^t) \bigodot h^{t-1} u^t + u^t \bigodot \tilde{h}^t \tag{1}$$

$$\tilde{h}^t = tanh(F_o([h^{t-1} \bigodot r^t, x^t])) \tag{2}$$

$$u^t = \sigma(F_u([h^{t-1}, x^t])) \tag{3}$$

$$r^t = \sigma(F_r([h^{t-1}, x^t])) \tag{4}$$

Among them, $\bigodot$ denotes pointwise multiplication and $[\cdot,\cdot]$ denotes vector connection. $F_o(\cdot)$, $F_u(\cdot)$, $andF_r(\cdot)$ denotes the sub-network of output, update gate, and reset gate, respectively. $\sigma(\cdot)$ denotes *sigmoid* activation function, and $tanh(\cdot)$ denotes *tanh* activation function. The *sigmoid* activation function maps the input to [0,1] to realize the function of the gate. The *tanh* activation function is used to adjust the range of alternative states to avoid gradient explosion.

Training Strategy

The proposed HDN adopts a step-by-step training strategy. Firstly, the short-term feature extractor SENet is trained supervised. The SENet is divided into two parts, the last two fully connected layers form a short-term classifier f_1, and the front-end part forms a short-term feature extractor g_1. Using the labeled short-term EEG samples to train g_1 and f_1, with cross-entropy loss function, the optimization target is as follows:

$$\min_{\Theta_{g_1},\Theta_{f_1}} \frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{M} y_{i,c} \log(\hat{y}_{i,c}) \tag{5}$$

Then, the long-term feature extractor LSTM is trained supervised. The short-term feature vector sequence output by g_1 is used as the input of LSTM. When training LSTM, the parameters of g_1 are fixed and no longer updated. Use g_2 to denote the parameters of the LSTM and f_2 to denote the parameters of the seizure classifier MLP. The cross-entropy loss function is also used. The optimization target is as follows:

$$\min_{\Theta_{g_2},\Theta_{f_2}} \frac{1}{N_b} \sum_{i=1}^{N_b} \sum_{c=1}^{M} y_{ic} \log(\hat{y}_{ic}) \tag{6}$$

Finally, the LSTM is fine-tuned *via* adversarial domain adaptation to realize domain-invariant deep representation. As shown in **Figure 2**, HDN, seizure classifier f_2, and domain discriminator f_d form an adversarial learning network. Among them, f_2 and f_d both take the deep features output by HDN as input, and both are MLPs containing two hidden layers. In the domain adaptive learning phase, there are two inputs and two outputs. The model uses two datasets for train: one is the EEG dataset that has been labeled, and the other is the unlabeled EEG sample from target patient. The domain discriminator f_d and HDN are optimized alternately. First, we freeze HDN and seizure classifier, and train f_d. Because the domain index is known, the binary cross-entropy loss function is used. The optimization goal is shown as follows:

$$\min_{\Theta_{f_d}} \frac{1}{N_b} \sum_{i=1}^{N_b} d_i \log(\hat{d}_i) + (1-d_i)\log(1-\hat{d}_i) \tag{7}$$

After the domain classification loss converges, we freeze the parameters of f_d, fine-tune the LSTM, and reduce the distribution discrepancy between the target domain and the source domain in the deep feature space. At the same time, the seizure classifier should be fine-tuned to maintain the classification accuracy in the

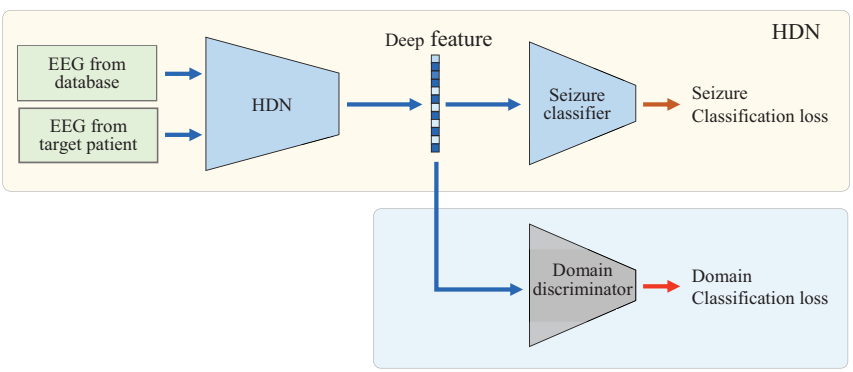


FIGURE 2 | Domain adaptation of the proposed HDN based on adversarial learning.

source domain. Therefore, at this phase, the optimization target consists of two sub-items, as shown in Eq. 8.

$$\min_{\Theta_{g_2},\Theta_{f_2}} \frac{1}{N_b}\sum_{i=1}^{N_b}\sum_{c=1}^{M} y_{ic}\log(\hat{y}_{ic}) - \frac{1}{N_b}\sum_{i=1}^{N_b} d_i\log(\hat{d}_i) + (1-d_i)\log(1-\hat{d}_i) \quad (8)$$

The EEG classifier learns the spatial distribution of deep features under the supervision of labeled samples and optimizes the classifier. By minimizing the classification loss, the classifier can classify the labeled samples accurately. By minimizing the loss of deep distribution differences, the encoder can extract deep features from samples of different ages and different body mass index (BMI) that obey the same distribution. In this way, when EEG monitoring equipment is applied to people of different ages and BMI, it is not necessary to label the EEG of the target population manually. The unlabeled EEG can drive the model to adjust adaptively and realize the personalized arrhythmia monitoring based on EEG.

RESULTS

EEG Database

The multi-class seizure type classification was implemented using the version 1.5.2 of Temple University Hospital (TUH) EEG Seizure Corpus (Roy et al., 2019). The train set of TUH EEG Seizure Corpus v1.5.2 collected 592 patient EEGs and recorded a total of 2,377 seizures. A total of eight types of seizure were recorded, of which three types had more than 50 seizures. The number of occurrences of other types is too less to train the deep model; they are not considered in this study. Therefore, in this study, the TUH database is used to carry out four classification experiments, which are Normal, Focal Non-Specific Seizure (FNSZ), Generalized Non-Specific Seizure (GNSZ), and Complex Partial Seizure (CPSZ).

The CHB-MIT database (Goldberger et al., 2000; Shoeb and Guttag, 2010) was also used in this research to verify the generalization of the proposed method when the test samples and training samples come from different hospitals and patients. The CHB-MIT database comes from 23 pediatric patients at Boston Children's Hospital and an adult patient at Beth Israel Deaconess Medical Center. A total of 173 seizures judged by experts were recorded. The EEG is digitized at a sampling frequency of 256 Hz and divided into 1-h records for storage. After review by experts, the onset and end time of each seizure were recorded.

Performance of Hybrid Deep Networks to Classify Seizure of EEG From Same Database

In this section, we use the TUH database to verify the classification accuracy of the proposed HDN after supervised training. We clip two samples from the EEG record of each seizure, one of which is centered on the midpoint of the seizure, and the other is centered on the starting point of the seizure. A total of 3,000 normal samples, 3,072 FNSZ samples, 818 GNSZ samples, and 566 CPSZ samples were clipped. The samples were randomly divided into six parts, of which six were used for training and one was used for testing, and a sixfold crossover experiment was performed. The Adam optimizer is utilized. The batch size is set to 128. The learning rate is set to 0.001. **Figure 3** shows the change curve of accuracy and loss in the second training phase. The accuracy rate reaches 85% within 50 epochs and then rises slowly and steadily. At the 700th epoch, it reached 95% and stabilized. After the first phase of supervised learning, SENet achieved an accuracy rate of over 80%. Therefore, the accuracy of HDN can be increased quickly in the second training phase.

Generalization of the Proposed Method to Classify Seizure for New Patients

In this section, we test the proposed method on different seizure classification tasks. We use different databases to train HDN and then test on different databases. The CHB-MIT database has only two types of labels, normal and seizure. Correspondingly, when using the TUH database to train the model, the samples are also relabeled as normal or seizure. **Table 1** lists the test accuracy of using different methods to train HDN in each task. In each task, the training set is further divided into a training set and a validation set, and a sixfold crossover experiment is carried out to optimize the hyper-parameters and then tested on the test set. In order to comprehensively verify the superiority of the proposed method, we also trained the HDN using samples in the test set, and list the validation accuracy. Obviously, when the test set and training set are the same, HDN achieves the highest classification accuracy.

The fourth column of **Table 1** lists the training accuracy achieved by HDN using each database for supervised learning. Among them, the CHB-MIT set is a normal/seizure two classification, and HDN achieves the highest training accuracy rate, reaching 97.8%. The training accuracy achieved in the TUH

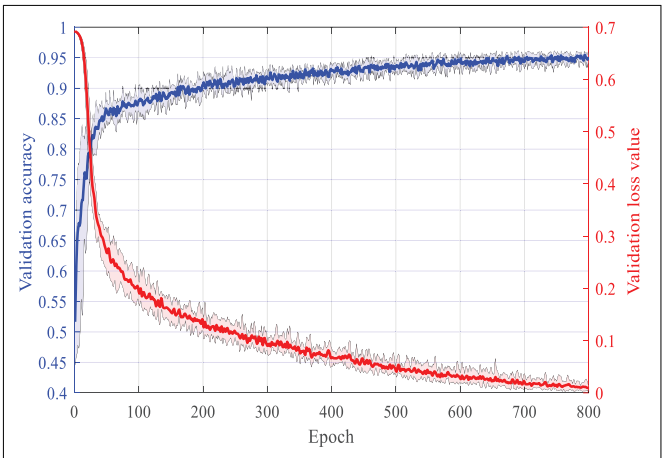


FIGURE 3 | The learning process of HDN under the supervision of the expert label.

TABLE 1 | The classification accuracy of hybrid deep networks (HDN) in the target domain trained by different method.

Train set	Test set	Trained with only train set	Trained with only test set	The proposed method
TUH train set	TUH dev set	92.4%	94.2%	94.7%
TUH dev set	TUH train set	71.2%	95.3%	80.3%
TUH train set	CHB-MIT set	89.7%	97.8%	96.3%
TUH dev set	CHB-MIT set	85.5%	97.8%	92.1%
CHB-MIT set	TUH dev set	80.3%	94.2%	86.7%

train set is the second, 95.3%. The training accuracy achieved in TUH dev set is the lowest at 94.2%. The author infers that this is because the same is a multi-classification task, but the TUH dev set samples are fewer.

Comparing the third and fourth columns, you can find the degradation of the trained HDN when applied to different test sets. The HDN trained with the TUH train set achieves a test accuracy of 92.4% in the TUH dev set, which is 2.9% lower than the training accuracy achieved by the HDN in the TUH train set, and is more accurate than the training achieved by the HDN in the TUH dev set. The rate dropped by 1.8%. This shows that under the condition of sufficient training samples, the recommended HDN has quite superior generalization ability. The high test accuracy of 89.7% achieved by the HDN trained with the TUH train set on the CHB-MIT set further validates this point. The HDN trained with the TUH dev set has the lowest test accuracy in the TUH train set, which is only 71.2%. Although HDN has achieved a training accuracy rate of over 94% in the TUH dev set, in the face of a larger and more complex TUH train set, the accuracy rate has dropped by more than 20%, losing its clinical application value. This verifies the importance of accumulating training data for automatic EEG analysis and diagnosis.

Finally, we compare column five and column three to verify the advantage of the proposed method. Following the proposed method, HDN first conducts supervised learning based on the training set. Then, we use the test set to carry out transfer learning to construct domain-invariant deep representation, thereby improving the adaptability of HDN in the test set. In task 1, the proposed method achieves a test accuracy of 94.7%, which exceeds the training accuracy achieved by HDN in the TUH dev set. When the training set is TUH dev set and the test set is TUH train set, the performance improvement of the proposed training method is the most significant, and the test accuracy rate is increased by 9.1%. When the test set and training set come from different patient groups, the proposed method increases the accuracy of HDN testing by more than 6 percentage points.

DISCUSSION

In this paper, we proposed a seizure classification method based domain-invariant deep representation of EEG. The proposed method is implemented by the HDN responsible for feature extraction and classification and the domain discriminator responsible for fine-tuning the deep representation. Benefiting from the advantages of the convolution kernel in extracting waveform features, after supervised learning, the CNN extracts deep features from the short-term EEG. The SE-Res module enables the modeling of the internal relationship of each lead, which improves the multi-classification capability of SENet. In the end, a seizure classification accuracy rate of 85.7 was achieved on the TUH database, which is a significant improvement compared to the machine learning model, but still cannot meet the needs of clinical applications. The author infers that this is because of the heterogeneity of seizure's representation in EEG and the interference of complex and diverse non-stationary noise. The separability of short-term EEG samples in the deep feature space is insufficient, and it is difficult for a MLP to optimize an accurate classification hyperplane. CNNs can process EEG samples for a longer period of time, but require a huge amount of calculation, which conflicts with the limited computing power of monitoring equipment. Therefore, this article uses the trained SENet as a short-term deep feature extractor, and then uses an LSTM that is good at processing time series to extract long-term deep features. SENet, LSTM, and NN classifiers together form HDN. After stepwise supervised learning, a classification accuracy rate of more than 92% is finally achieved on the TUH database. The accuracy of detecting seizure on the CHB-MIT database reached 97.8%. The objective reasons include two points. One is that the CHB-MIT database contains far fewer patients than the TUH database, and the other is that the normal/seizure two-class classification task is more important and simple. However, under different combinations of training set and test set, the test accuracy of HDN has dropped by 3–13 percentage points compared to the evaluation accuracy. This shows that the generalization ability of HDN is still insufficient. What is exciting is that adversarial learning based on labeled data in the training set and unlabeled data in the test set can drive HDN to fine-tune the deep feature representation, so that the test set samples and training set samples are aligned in the deep space, thereby improving the test accuracy of the classifier. In task 2 where the test set is larger than the training set, the fine-tuning re-representation increased the test accuracy by 9.1 percent to 80.3%. When the test set and training set come from different hospitals and patient groups, the fine-tuning re-representation increased the test accuracy by more than 5 percentage points. In summary, the constructed HDN has strong EEG feature extraction capabilities, and the proposed adversarial learning method can adjust the deep feature mapping to achieve excellent domain-independent properties, which has important value for the clinical application of seizure automatic classification equipment.

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 drafted the manuscript and performed all administrative tasks required for submission. BY took part in planning, supervision, and brainstorming the manuscript. BY and BC designed the classification method and contributed to the interpretation of results. WS contributed to the statistical analysis and interpretation of results. GT contributed to the identification of research topic and preparation of study design. All authors critically revised the manuscript and approved the final version.

FUNDING

This research was supported by the National Key Research and Development Program of China (No. 2020YFB1713503) and the Fundamental Research Funds for the Central Universities (No. 20720190009).

REFERENCES

Abdelhameed, A., and Bayoumi, M. (2021). A deep learning approach for automatic seizure detection in children with epilepsy. *Front. Comput. Neurosci.* 15:650050. doi: 10.3389/fncom.2021.650050

Aileni, R. M., Pasca, S., and Florescu, A. (2019). "Epileptic seizure classification based on supervised learning models," in *Proceedings of the 2019 11th International Symposium on Advanced Topics in Electrical Engineering, In International Symposium on Advanced Topics in Electrical Engineering*, Bucharest.

Alghamdi, A., Hammad, M., Ugail, H., Abdel-Raheem, A., and Aaa El-Latif, A. A. (2020). Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *Multimedia Tools Applic.* 1–22. doi: 10.1007/978-3-030-15145-4_65-2

Arsenovic, M., Sladojevic, S., Gavrilov, K., Anderla, A., and Sladojevic, M. (2020). "Application of deep learning for automated arrhythmia detection using ECG signals," in *Proceedings of the 18th International Scientific Conference on Industrial Systems Industrial Innovation in Digital Age*, Novi Sad.

Cao, X., Chen, B., and Zeng, N. (2020). A deep domain adaption model with multi-task networks for planetary gearbox fault diagnosis. *Neurocomputing* 409, 173–190. doi: 10.1016/j.neucom.2020.05.064

Cao, X., Wang, Y., Chen, B., and Zeng, N. (2021). Domain-adaptive intelligence for fault diagnosis based on deep transfer learning from scientific test rigs to industrial applications. *Neural Comput. Applic.* 33, 4483–4499. doi: 10.1007/s00521-020-05275-x

Chen, B., He, W., Hu, J., Wang, G., and Guo, B. (2019). Method for diagnosis of data-driven GMC sparse enhancement GMC. *J. Xidian Univ.* 46, 74–79. doi: 10.19665/j.issn1001-2400.2019.04.011

Gabara, A., Yousri, R., Hamdy, D., Zakhari, M. H., and Mostafa, H. (2020). "Patient specific epileptic seizures prediction based on support vector machine," in *Proceedings of the 2020 32nd International Conference on Microelectronics, In International Conference on Microelectronics-ICM*, Aqaba, 40–43.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation [Online]* 101, e215–e220. doi: 10.1161/01.cir.101.23.e215

He, W., Hu, J., Chen, B., Li, C., and Guo, B. (2019). Sparsity-induced resonance demodulation method for blade crack detection. *J. Xidian Univ.* 46, 75–80. doi: 10.19665/j.issn1001-2400.2019.06.011

He, W., Sun, W., Su, B., Yang, Y., and Guo, B. (2018). Sparse feature extraction technique and its applications to machinery fault diagnosis. *J. Xidian Univ.* 45, 154–159. doi: 10.3969/j.issn.1001-2400.2018.02.026

Hussain, W., Wang, B., Niu, Y., Gao, Y., Wang, X., Sun, J., et al. (2019). Epileptic seizure detection with permutation fuzzy entropy using robust machine learning techniques. *IEEE Access* 7, 182238–182258. doi: 10.1109/access.2019.2956865

Krumholz, A., Wiebe, S., Gronseth, G., Shinnar, S., Levisohn, P., Ting, T., et al. (2007). Practice Parameter: evaluating an apparent unprovoked first seizure in adults (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology and the American Epilepsy Society. *Neurology* 69, 1996–2007. doi: 10.1212/01.wnl.0000285084.93652.43

Li, M., and Chen, W. (2021). FFT-based deep feature learning method for EEG classification. *Biomed. Signal. Process. Control* 66:102492. doi: 10.1016/j.bspc.2021.102492

Li, Y., Yu, Z., Chen, Y., Yang, C., Li, Y., Allen Li, X., et al. (2020b). Automatic seizure detection using fully convolutional nested LSTM. *Int. J. Neural Syst.* 30:2050019. doi: 10.1142/s0129065720500197

Li, Y., Liu, Y., Cui, W.-G., Guo, Y.-Z., Huang, H., and Hu, Z.-Y. (2020a). Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 782–794. doi: 10.1109/tnsre.2020.2973434

Liu, G., Zhou, W., and Geng, M. (2019). Automatic seizure detection based on s-transform and deep convolutional neural network. *Int. J. Neural Syst.* 30:1950024. doi: 10.1142/s0129065719500242

Mahjoub, C., Le Bouquin Jeannes, R., Lajnef, T., and Kachouri, A. (2020). Epileptic seizure detection on EEG signals using machine learning techniques and advanced preprocessing methods. *Biomed. Eng. Biomed. Technik* 65, 33–50. doi: 10.1515/bmt-2019-0001

Naz, M., Shah, J. H., Khan, M. A., Sharif, M., and Damasevicius, R. (2021). From ECG signals to images: a transformation based approach for deep learning. *PeerJ Comput. Sci.* 7:e386. doi: 10.7717/peerj-cs.386

Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., and Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv* Preprint. arXiv:1707.01836.

Roy, S., Asif, U., Tang, J., and Harrer, S. (2019). "Seizure type classification using EEG signals and machine learning: setting a benchmark," in *Proceedings of the 2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Philadelphia, PA.

Shabarinath, B. B., Challagulla, K., and Visodhan, M. R. (2019). "A comparative study of epileptic seizure detection framework using SVM and ELM," in *Proceedings of the 2019 International Conference on Intelligent Computing and Control Systems*, Madurai.

Shoeb, A. H., and Guttag, J. V. (2010). "Application of machine learning to epileptic seizure detection," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa.

Slimen, I. B., Boubchir, L., Mbarki, Z., and Seddik, H. (2020). EEG epileptic seizure detection and classification based on dual-tree complex wavelet transform and machine learning algorithms. *J. Biomed. Res.* 34, 151–161. doi: 10.7555/jbr.34.20190026

Subasi, A., Kevric, J., and Abdullah Canbaz, M. (2019). Epileptic seizure detection using hybrid machine learning methods. *Neural Comput. Applic.* 31, 317–325. doi: 10.1007/s00521-017-3003-y

Wang, H., He, H., and Katabi, D. (2020). "Continuously indexed domain adaptation," in *Proceedings of the 37th International Conference on Machine Learning, ICML 2020 37th International Conference on Machine Learning*, eds H. Daume and A. Singh, International Machine Learning Society (IMLS), 9840–9849.

Wang, S.-H., Xie, S., Chen, X., Guttery, D. S., Tang, C., Sun, J., et al. (2019). Alcoholism identification based on an AlexNet transfer

learning model. *Front. Psychiatry* 10:205. doi: 10.3389/fpsyt.2019.00205

Wang, Y., Cao, J., Wang, J., Hu, D., and Deng, M. (2019). "Epileptic signal classification with deep transfer learning feature on mean amplitude spectrum," in *Proceedings of the 2019 41st Annual International Conference of the Ieee Engineering in Medicine and Biology Society*, In IEEE Engineering in Medicine and Biology Society Conference Proceedings, (New York, NY: IEEE), 2392–2395.

World Health Organization (2019). Available online at: https://www.who.int/publications/i/item/epilepsy-a-public-health-imperative

Yıldırım, Ö, Pławiak, P., Tan, R. S., and Acharya, U. R. (2018). Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput. Biol. Med.* 102, 411–420. doi: 10.1016/j.compbiomed.2018.09.009

Yildirim, O., Baloglu, U. B., and Acharya, U. R. (2020). A deep convolutional neural network model for automated identification of abnormal EEG signals. *Neural Comput. Applic.* 32, 15857–15868. doi: 10.1007/s00521-018-3889-z

Yin, W., Yang, X., Li, L., Zhang, L., Kitsuwan, N., Shinkuma, R., et al. (2019). Self-adjustable domain adaptation in personalized ECG monitoring integrated with IR-UWB radar. *Biomed. Signal Process. Control* 47, 75–87. doi: 10.1016/j.bspc.2018.08.002

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2021). A comprehensive survey on transfer learning. *Proc. IEEE* 109, 43–76. doi: 10.1109/jproc.2020.3004555

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fnins.2021.760979

A Generalizable and Discriminative Learning Method for Deep EEG-Based Motor Imagery Classification

Xiuyu Huang[1], Nan Zhou[2,1] and Kup-Sze Choi[1]**

[1] Centre for Smart Health, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [2] College of Control Engineering, Chengdu University of Information Technology, Chengdu, China

Convolutional neural networks (CNNs) have been widely applied to the motor imagery (MI) classification field, significantly improving the state-of-the-art (SoA) performance in terms of classification accuracy. Although innovative model structures are thoroughly explored, little attention was drawn toward the objective function. In most of the available CNNs in the MI area, the standard cross-entropy loss is usually performed as the objective function, which only ensures deep feature separability. Corresponding to the limitation of current objective functions, a new loss function with a combination of smoothed cross-entropy (with label smoothing) and center loss is proposed as the supervision signal for the model in the MI recognition task. Specifically, the smoothed cross-entropy is calculated by the entropy between the predicted labels and the one-hot hard labels regularized by a noise of uniform distribution. The center loss learns a deep feature center for each class and minimizes the distance between deep features and their corresponding centers. The proposed loss tries to optimize the model in two learning objectives, preventing overconfident predictions and increasing deep feature discriminative capacity (interclass separability and intraclass invariant), which guarantee the effectiveness of MI recognition models. We conduct extensive experiments on two well-known benchmarks (BCI competition IV-2a and IV-2b) to evaluate our method. The result indicates that the proposed approach achieves better performance than other SoA models on both datasets. The proposed learning scheme offers a more robust optimization for the CNN model in the MI classification task, simultaneously decreasing the risk of overfitting and increasing the discriminative power of deeply learned features.

Keywords: electroencephalogram, motor imagery, convolutional neural networks, label smoothing, center loss

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Hongru Zhao,
Soochow University, China
Yufeng Yao,
Changshu Institute of Technology, China

****Correspondence:***
Xiuyu Huang
xiuyu.huang@connect.polyu.hk
Kup-Sze Choi
thomasks.choi@polyu.edu.hk

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *19 August 2021*
Accepted: *06 September 2021*
Published: *22 October 2021*

Citation:
Huang X, Zhou N and Choi K-S (2021) A Generalizable and Discriminative Learning Method for Deep EEG-Based Motor Imagery Classification. Front. Neurosci. 15:760979. doi: 10.3389/fnins.2021.760979

1. INTRODUCTION

Brain–computer interface (BCI) has been raising interest from the research community. It provides an important way for the disabled to interact with the outside world without using any muscular movements (Wolpaw et al., 2002). This technology aims to recognize the user intentions based on the distinct patterns of neural events. Motor imagery (MI) is one of the crucial topics in the area of BCI, referring to a cognitive procedure of the motion imagination such as lifting left or right leg, without any actual moving actions (Ahn and Jun, 2015). The most popular technology to signalize such cognitive procedures is the electroencephalogram (EEG), being noninvasive and

relatively easy to set up (Ahn and Jun, 2015; Ni et al., 2020; Zhang et al., 2020). The principle of the EEG-based MI-BCI system is to match the type of motion imagination and its corresponding EEG signals. Such matching systems have been practiced in a variety of applications, including speller (Rezeika et al., 2018), wheelchair (Kaufmann et al., 2014), and prosthesis (Vidaurre et al., 2016).

Accurate classification of EEG-MI pattern is one of the most decisive factors to the BCI performance but remains a significant challenge due to the low signal-to-noise ratio (SNR) characteristics of the EEG signal (Goldenholz et al., 2009; Zhang et al., 2019). Convolutional neural networks (CNNs) have been widely explored and achieved great success in the MI recognition area (Bashivan et al., 2015; Roy et al., 2019). It significantly pushes the boundary of the state-of-the-art (SoA) in classification accuracy compared to the conventional methods such as band power analysis (Martinez-Leon et al., 2015), independent component analysis (ICA) (Lee et al., 1999), and common spatial filter (CSP) (Ramoser et al., 2000). The most common framework of CNN is to perform feature generation and label prediction, learning deep features from raw EEG data by the CNN pipeline, then making label predictions based on the learned features (see **Figure 1A**).

The training of CNNs in the MI classification task is mainly guided by minimizing the cross-entropy (Hertz et al., 2018). This objective function is "greedy" and encourages the largest possible logit gaps, making the model less adaptive, and sometimes overconfident to its predictions (Szegedy et al., 2016). The model learns to assign a full probability to the ground-truth label for each training example, even though some noisy data are mixed in the training set (Müller et al., 2019). This overfitting phenomenon incredibly easily occurs when the training sample size is small. Coincidentally, the EEG data have a low SNR and contains much noise. In addition, the MI-BCI system is usually designated as subject dependent, so it usually has limited training data. To reduce the risk of overfitting (overconfidence) issue, we adopt a label smoothing technique introduced by Szegedy et al. (2016) in the training scheme of the MI classification. It computes a modified cross-entropy, called smoothed cross-entropy, not by using "hard" one-hot encoded labels such as [0, 0, 1, 0] from training data, but a weighted mixture of these hard labels with the uniform distribution (**Figure 1B**). Label smoothing alternatively encourages small logit gaps and prevent overconfident predictions. This technology has successfully increased the performance of CNN models across multiple tasks, including image classification (Szegedy et al., 2016), speech recognition (Chorowski and Jaitly, 2016), and machine translation (Vaswani et al., 2017). It is expected to benefit the model training in MI classification by tackling the overfitting problem, leading to a generalizable and adaptive CNN model.

In addition to ensuring the model's generalizability, we also aim to increase its discriminative power. As shown in the common framework of CNN (**Figure 1A**), the last fully connected layer acts as a linear classifier, and the cross-entropy only encourages the separability (Hertz et al., 2018) but does not guarantee the high discriminative characteristics, where features have both a large inter-class difference and a tight intra-class variation (**Figure 1C**). Therefore, the resulting features generated by the model trained via the cross-entropy are not sufficiently effective for the MI classification. To enhance the discriminative capacity of the deep features, we apply a center loss (Wen et al., 2016) for the model training. Specifically, the center of deep features is computed by their means in each class and updated across every epoch. The distances between deep features with their corresponding class centers are minimized at each training iteration. The parameters are optimized by jointly minimizing the cross-entropy and center loss. Intuitively, the cross-entropy forces the deep features from different classes to stay apart, and the center loss pulls the features belonging to the same class toward their centers. With joint supervision, we can concurrently enlarge the inter-class difference and reduce the intra-class variation so as to improve the discriminative power of deep features.

In this paper, we propose a novel training scheme for CNN-based model in the MI classification by using a combined loss with smoothed cross-entropy and center loss. The main contributions are as follows:

1. To our best knowledge, although structures of the CNN model are heavily investigated, this is the first attempt to use the proposed loss to help supervise training in the context of MI classification. With joint supervision of the smoothed cross-entropy and the center loss, both generalizable and discriminative model can be obtained for robust MI recognition.
2. We present extensive experiments on two famous MI public datasets, called BCI-competition IV-2a and IV-2b. Our new approach achieves superior performance compared to other SoA methods.
3. We also conduct an ablation study to demonstrate the effectiveness of the label smoothing and the center loss.

The remainder of the paper is organized as follows. In section 2, conventional and deep learning methods on MI classification are introduced. Section 3 describes the proposed approach. Sections 4 and 5 present the experiment result and analysis. Section 6 concludes the current study.

2. RELATED WORKS

A sophisticated feature extractor is the key to success in conventional methods for the MI classification task. One of the most frequently and widely used approaches is the common spatial pattern (CSP) (Pfurtscheller and Neuper, 2001; Yu et al., 2019). It tries to generate optimal spatial filters that have minimum or maximum variance between classes in a particular frequency band. The features used in the winner algorithm of the BCI competition IV are based on the filter bank CSP (FBCSP) (Ang et al., 2008) that finds a set of optimal spatial filters in multiple frequency bands. The Naive Bayes Parzen Window classifier using these features achieved an outstanding classification performance with an accuracy of 67.75% on the dataset IV-2a (Ang et al., 2008). After the competition, a novel method based on the support vector machine (SVM) with Riemannian covariance achieved a better performance (75.74%)

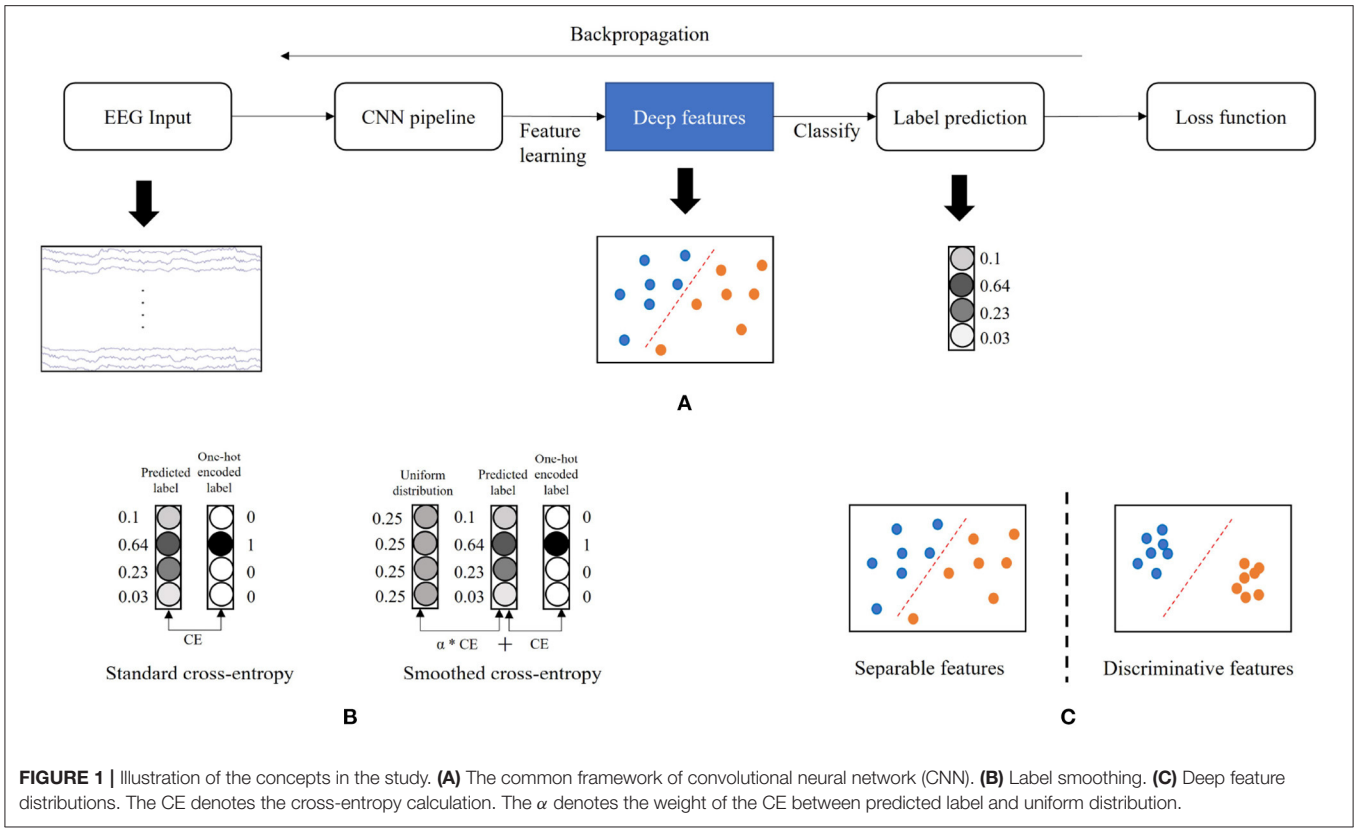


FIGURE 1 | Illustration of the concepts in the study. **(A)** The common framework of convolutional neural network (CNN). **(B)** Label smoothing. **(C)** Deep feature distributions. The CE denotes the cross-entropy calculation. The α denotes the weight of the CE between predicted label and uniform distribution.

for the same database (Hersche et al., 2018). In addition to these vector-based methods, matrix-form strategies such as the logistics regression classifier with a nuclear norm regularization (Zhou and Li, 2014), the rank-k SVM (Lal et al., 2004), and the support matrix machine (SMM) (Zheng et al., 2018) were also developed by multiple research groups. The leading edge of these methods is to directly process the 2-D MI EEG data on a matrix basis instead of stacking features as a vector input to a classifier, which preserves the informative structural patterns.

Deep learning (DL) models were also exploited to tackle the MI classification challenge. For instance, the multilayer perceptron (MLP) was proposed to generate nonlinear patterns from CSP features and also to substitute the SVM as a classifier for MI recognition (Kumar et al., 2016). Similarly, a channel-wise convolution with channel mixing (C2CM) was introduced to classify the spatial-temporal features generated by FBCSP (Sakhavi et al., 2018). Bashivan et al. (2015) converted the EEG waves into spectral topographies via short-time Fourier transform (STFT). These topographies were then fitted into CNNs for further transformation and classification. Tabar and Halici (2016) also used the STFT approach to extract spatial-temporal images as the feature input to the CNN-SAE model for classification. These feature input (FI) models still require complex feature generation from raw EEG data prior to the DL modeling. Several research groups investigate raw signal input (RSI) models to provide an end-to-end scheme for MI recognition to address this limitation. For example, the two most well-known RSI networks, EEGNet (Lawhern et al., 2018), and ConvNet (Schirrmeister et al., 2017), achieved competitive classification performance without using any pre-processing techniques. In general, although model architectures were heavily investigated, the neural network learning process did not receive too much attention from the MI community. Rather than figuring out a more sophisticated architecture, we propose a potentially efficient objective function for both generalizable and discriminative learning in the CNN-based model.

3. METHOD

This section first introduces the notations and definitions used in this work and describes the CNN architecture. Then, the novel proposed loss is presented in detail.

3.1. Definition and Notations

Assuming that the DL model input is on a per-trial basis, where the continuous EEG is segmented into labeled trials, we define the segmented trials of a subject as $\{[x_i, y_i]\}_{i=1}^{n}$, where $x_i \in \mathbb{R}^{E\times T}$ represents *ith* of EEG trials recorded by E electrodes and T sampling time points. $y_i \in \mathbb{R}^{M}$ denotes the corresponding i^{th} labels of M classes. Let the ground truth distribution p over labels $p(y|x_i)$, and $\sum_{y=1}^{M} p(y|x_i) = 1$. We also define a CNN-based model with θ that predicts label distribution $q_\theta(y|x_i)$, and certainly $\sum_{y=1}^{M} q_\theta(y|x_i) = 1$. We are

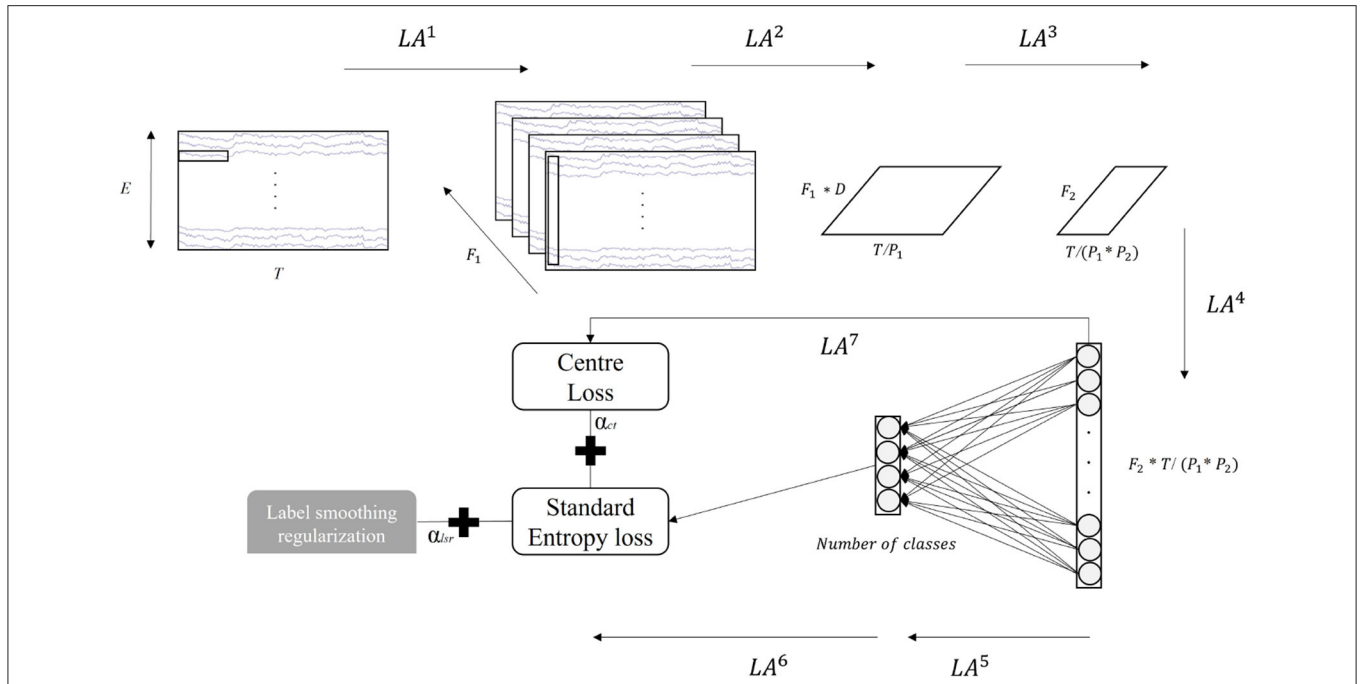


FIGURE 2 | Illustration of the convolutional neural network (CNN)-based model. T = number of the timestamps, E = number of electricodes, F_1 = number of filters in the temporal CNN, D = number of depthwise convolution output channels, F_2 = number of spatial filters, P_1 and P_2 are sizes of the average pooling kernel, and α_{lsr} and α_{ct} are weights of label smoothing regularization and center loss.

motivated to adopt the label smoothing technology and center loss to improve the generalizable and discriminative power of the CNN-based model.

3.2. Network Architecture

In the current study, we inherit the CNN architecture of the EEGNet (Lawhern et al., 2018) but make two modifications, where the kernel size of the first temporal CNN filter (LA^1) is decreased to attain temporal information above 8 Hz, as the alpha (8–12 Hz) and beta (12.5–30 Hz) band contain most relevant information of the motor imagery task (Wierzgała et al., 2018). The illustration of the network is displayed in **Figure 2**, and the details of each layer are presented in **Table 1**. The model begins with a 2D-CNN directly linked to the raw EEG data with a kernel size in (K_1, 1) to capture temporal patterns in each electrode. A depthwise convolution layer with a kernel size of (1, E) is followed and utilized for spatial feature extraction. The separableConv2D with a kernel size in (K_2,1) is then performed to gain deeper and more abstract temporal information across all electrodes. As shown in **Table 1**, it is noted that batch normalization (Ioffe and Szegedy, 2015), exponential linear unit (ELU) (Clevert et al., 2015) activation, and average pooling are sequentially followed after some of these convolutions for covariate shift avoidance (Bickel et al., 2009), nonlinear transformation, and dimension reduction, respectively. The deep feature generated by the CNN pipeline is then flattened as a vector (nodes) by a flatten layer. The vectors of each training batch are used to compute the center loss. The dense layer is subsequently connected to these nodes and acts as a classifier. The *softmax* function finally performs the estimation of the probability for each MI class. The cross-entropy between the probability estimation and the smoothed label represents the classification loss (standard cross-entropy + label smoothing regularization). The weighted sum of the classification and center losses supervises the training of the entire network.

3.3. Proposed Loss

To ensure both generability and discriminative power of the CNN-based model in the context of MI classification, we propose a combined loss that jointly optimizes the classification loss (e.g., cross-entropy computed by smoothed labels) and the center loss. Most models in previous studies performed the standard cross-entropy for the objective function defined as

$$\begin{aligned} L_{cl} &= \sum_{i=1}^{n} H_i(p, q_\theta) \\ &= -\sum_{i=1}^{n}\sum_{y=1}^{M} p(y|x_i) log\ q_\theta(y|x_i) \end{aligned} \tag{1}$$

Given the $p(y|x_i)$ is one-hot encoded in classification task where

$$p(y|x_i) = \begin{cases} 1, & y = y_i; \\ 0, & otherwise. \end{cases} \tag{2}$$

TABLE 1 | Architecture setting.

Layer	Function	Filter	Kernel	Output shape
	Input			$T, E, 1$
LA^1	Conv2d	F_1	$(K_1, 1)$	T, C, F_1
	BatchNorm			
	DepthwiseConv2d	$F_1 * D$	$(1, E)$	
	BatchNorm			$T, 1, F_1 * D$
LA^2	ELU Activation			
	Average pooling		$(P_1, 1)$	$T/P_1, 1, F_1 * D$
	SparableConv2d	F_2	$(K_2, 1)$	
	BatchNorm			$T/P_1, 1, F_2$
LA^3	ELU Activation			
	Average pooling		$(P_2, 1)$	$T/(P_1 * P_2), 1, F_2$
LA^4	Flatten			$(T * F_2)/(P_1 * P_2)$
LA^5	Fully connected			*number of classes*
LA^6	Softmax (CEL)			
LA^7	Lambda (CL)			1

(1) T = number of the timestamps, E = number of electricodes, K_1 = kernel size of the first CNN, D = number of depthwise convolution output channels, F_1 = number of temporal filters, F_2 = number of spatial filters, K_2 = size of th kernel in the spatial filer, and P_1 and P_2 are sizes of average pooling kernels.

(2) CEL stands for cross-entropy computed by smoothed labels. CL stands for center loss. Lamda layer is a self-customize layer for the calculation of the center loss.

We can further reduce (1) as

$$L_{cl} = -\sum_{i=1}^{n} log\ q_\theta(y_i|x_i) \tag{3}$$

For each training sample i, the $q_\theta(y_i|x_i)$ is usually calculated by the *softmax* function as follows:

$$q_\theta(y_i|x_i) = \frac{exp(z_{y_i})}{\sum_{j=1}^{M} exp(z_j)} \tag{4}$$

Here, z_j is the logit value or unnormalized log-probability for each label j. By using the one-hot ground-truth label, minimizing the objective function L_{cl} is equivalent to do the log-likelihood maximum. The maximum is not achievable with finite data, so it can only be estimated in the case when $z_{y_i} >> z_j$ for all $j \neq y_i$ (e.g., the logit of the ground-truth label is much larger than all other logits) over the training dataset (Szegedy et al., 2016). In such a case, the model learns to classify every training sample correctly with a confidence of nearly 1, which is the signal of overfitting. This phenomenon relatively easily occurs in the scenario that the MI EEG task often only contains a small sample size of training data.

We adopt the label smoothing mechanism where a noise distribution $u(y|x)$ is added to the one-hot ground truth label to prevent the model from having overconfidence and to reduce the risk of overfitting. Then, the new ground truth label distribution is $p'(y|x_i) = (1-\varepsilon)p(y|x_i) + \varepsilon u(y|x_i)$, where ε is a weight factor,

TABLE 2 | Hyperparameter settings.

Hyperparameter	CNN for IV-2a	CNN for IV-2b
T	1,000	1,000
E	22	3
F_1	8	8
F_2	16	16
K_1	32	32
K_2	16	16
D	2	2
P_1	8	8
P_2	8	8

$\varepsilon \in [0, 1]$. By replacing $p(y|x_i)$ with $p'(y|x_i)$ in (1) and (2), the new classification loss L'_{cl} computed by smoothed labels can be written as

$$\begin{aligned} L'_{cl} &= \sum_{i=1}^{n}\sum_{y=1}^{M} p'(y|x_i) log\ q_\theta(y|x_i) \\ &= -\sum_{i=1}^{n}\sum_{y=1}^{M}[(1-\varepsilon)p(y|x_i) + \varepsilon u(y|x_i)]\ q_\theta(y|x_i) \\ &= (1-\varepsilon)\sum_{i=1}^{n} H_i(p, q_\theta) + \varepsilon \sum_{i=1}^{n} H_i(u, q_\theta) \end{aligned} \tag{5}$$

The first half of (5) is the weighted standard cross-entropy L_{cl}. Let the second half is weighted loss of label-smoothing regularization (L_{lsr}), which penalizes the deviation of predicted label distribution p from noise distribution u with a relative weight $\varepsilon/(1-\varepsilon)$. We set the noise as the uniform distribution $u(y|x) = 1/M$ (Szegedy et al., 2016). Then, the $H_i(u, q_\theta)$ measures the dissimilarity between predicted label distribution p to uniform. Therefore, L_{lsr} heavily penalizes overconfident predictions and prevents poor generalization during the training.

Let the weight of L_{cl} is fixed as 1, and the relative weight $\varepsilon/(1-\varepsilon)$ of L_{lsr} is redefined as α_{lsr}. Therefore, L'_{cl} can be elaborated as

$$L'_{cl} = L_{cl} + \alpha_{lsr} L_{lsr} \tag{6}$$

In addition to the generalizability, we also try to ensure the discriminative ability of the deep learned features extracted by the CNN pipeline. Intuitively, simultaneously maximizing the inter-class distance and minimizing the intra-class variation is the fundamental strategy to keep features of different classes divisible. The cross-entropy minimization only assures to enlarge the inter-class distance, so we further employ a center loss to achieve intra-class variation reduction. We follow the equation proposed by Wen et al. (2016), and the center loss is defined as

$$L_{ct} = \frac{1}{2}\sum_{i=1}^{n} ||x_i - c_{y_i}||_2^2 \tag{7}$$

where c_{y_i} denotes the y_i center of the feature extracted by the CNN pipeline. Minimizing the distance between each deeply learned feature and its class center naturally decreases the intra-class variation. Finally, the objective function L is to jointly optimize the classification loss L'_{cl} and L_{ct}, defined as

$$L = L'_{cl} + \alpha_{ct} L_{ct} \tag{8}$$

where α_{ct} is the weight of center loss. Based on (6), L can be re-defined as,

$$L = L_{cl} + \alpha_{lsr} L_{lsr} + \alpha_{ct} L_{ct} \tag{9}$$

4. EXPERIMENTS

The BCI competition IV-2a and IV-2b datasets are used to evaluate the proposed approach. These two datasets are publicly available. The people involved in the datasets have obtained ethic approval. Users can download the data for free for research and publish relevant articles, so the ethical review and approval were waived for the current study.

4.1. Dataset Description

The BCI competition IV-2a (Tangermann et al., 2012) was recorded from 9 healthy individuals (A01-A09) by 22 EEG and 3 EOG channels in a sample rate of 250 Hz. The cue-based paradigm is used during the data collection. It consists of four MI classes, including the imagination motions of the left hand, right hand, tongue, and both feet. Two separate sessions were implemented for each subject. Each session comprises a total of 4*72 (a single MI class) = 288 trials. For fair comparisons with other approaches, the same data division scheme as that in the competition was used in our experiment. The first section is for model training and the second for model testing. Only the 4-s temporal segment (from the start of the cue until the end of the MI) in each trial is used in our model. Given the 250 Hz sample rate, our experiment's training and testing data are on a 1,000-sample series basis.

The BCI competition IV-2b (Tangermann et al., 2012) was also collected from nine healthy people (B01–B09) but only with 3 EEG channels (C3, Cz, and C4) attached to the frontal cortex. The dataset comprises two MI classes, including left-hand and right-hand movement imagination based on a cue-based BCI paradigm. Five independent sessions were recorded for each individual. We also keep using the same data division as that in the competition. The first three sessions are for training, and the remaining two are for evaluation. The 4-s temporal interval, from the starting point of the cue until the end of the MI, is used as a trial in our experiment. Given the recording frequency of 250 Hz, each training or testing trial is also on a 1,000-point basis.

4.2. Experimental Setup

Our approach is performed on a Tesla V100-SXM2 GPU running on Google online platform (Colab). The CNN network and the proposed loss function are implemented by Keras. The model is trained with Adam (Kingma and Ba, 2014) optimizer using a learning rate of 0.001, mini-batch size of 64, and 750 epochs. According to the result of the ablation study stated in section 5.2, the loss weights α_{lsr} and α_{ct} are set as 0.5 and 0.5, respectively. Other hyperparameters of the model architecture are shown in **Table 2**.

To evaluate the effectiveness of the proposed approach, we compare our strategy against existing SoA methods, including two conventional approaches [a vector-based method, e.g., the competition winner algorithm FBCSP (Ang et al., 2008), and a matrix-based method, SMM (Zheng et al., 2018)], two compact well-known DNN methods [EEGNet (Lawhern et al., 2018) and shallow ConNet (Schirrmeister et al., 2017)], and one more DNN methods (DRDA, Zhao et al., 2020) with complex architecture. The evaluation metric is the classification accuracy (*acc*).

5. RESULTS

5.1. Comparison With State-Of-The-Art Methods

The comparisons between the proposed methods and other models on the BCI competition IV-2a and IV-2b datasets are shown in **Tables 3**, **4**, respectively. The classification accuracy of each subject and the average accuracy are reported in a subject-dependent basis (e.g., training and testing data are from the same subject) as the same as the competition data division scheme. The model that has the best performance for each subject is highlighted in boldface. **Tables 3**, **4** clearly show that the proposed strategy has the best classification accuracy for nearly all subjects on both datasets with a maximum of 14.16% (subject A05) better than the second-best on IV-2a and of 11.43% (B02) better than second-best on IV-2b. On the average level, the classification average accuracies of our method have improvements of around 5.33 and 3.54% on IV-2a and IV-2b compared to other SoAs. We conduct paired t-tests between our approach and other SoA strategies to verify if the improvements are statistically significant. The p values obtained from the tests are indicated in **Table 5**. We can see that all p values are <0.05, which advises that the performance improvements of our method against others are statistically significant. In addition, it also can be seen that the corresponding standard deviations (SDs) of our method are 10.07 and 8.54%, which are both the smallest SD on the respective datasets. This result suggests that our method is a more robust classifier in a subject-independent manner than other models. All these results, as mentioned earlier, demonstrate that the CNN model trained using the proposed loss provides a more accurate and stable classification outcome for the MI recognition task.

5.2. Ablation Result Analysis

Ablation studies are carried out to study the contributions of the label smoothing technique and center loss to the CNN modeling. The hyperparameter α_{lsr} controls the degree of the smoothness on the label, and α_{ct} dominates the intra-class variations of the deep features. They are both significant. Therefore, two experiments on dataset IV-2a are explored to investigate the sensitiveness of these two hyperparameters.

5.2.1. Label-Smoothing Regularization

In the first experiment, we fix the α_{ct} as 0, where no center loss is applied, and vary α_{lsr} from 0 to 1 (inclusive) to train different

TABLE 3 | Classification accuracies (%) obtained with the dataset BCI competition IV-2a.

Methods	Subject									Average (SD)
	A01	A02	A03	A04	A05	A06	A07	A08	A09	
FBCSP	76.00	56.50	81.25	61.00	55.00	45.25	82.75	81.25	70.75	67.75 (12.94)
SMM	81.94	59.38	81.60	62.85	59.03	49.36	86.11	77.78	78.47	70.72 (12.35)
EEGNet	85.76	61.46	88.54	67.01	55.90	52.08	89.58	83.33	**86.81**	74.50 (14.36)
ConNet	76.39	55.21	89.24	74.65	56.94	54.17	**92.71**	77.08	76.39	72.53 (13.42)
DRDA	83.19	55.14	87.43	**75.28**	62.29	57.15	86.18	83.61	82.00	74.75 (12.22)
Ours	**89.32**	**66.78**	**94.14**	74.56	**76.45**	**62.33**	86.28	**85.61**	85.23	**80.08 (10.07)**

Highest values are highlighted in boldface.

TABLE 4 | Classification accuracies (%) obtained with the dataset BCI competition IV-2b.

Methods	Subject									Average (SD)
	B01	B02	B03	B04	B05	B06	B07	B08	B09	
FBCSP	70.00	60.36	60.94	**97.50**	93.12	80.63	78.13	92.50	86.88	80.01 (13.06)
SMM	67.81	51.79	53.44	93.31	82.81	74.69	72.19	82.50	75.62	72.68 (12.77)
EEGNet	68.44	57.86	61.25	90.63	80.94	63.13	84.38	93.13	83.13	75.88 (12.57)
ConNet	76.56	50.00	51.56	96.88	93.13	85.31	83.75	91.56	85.62	79.37 (16.27)
DRDA	81.37	62.86	63.63	95.94	93.56	88.19	85.00	95.25	**90.00**	83.98 (11.94)
Ours	**83.33**	**74.29**	**72.65**	96.09	**95.97**	**88.84**	**92.24**	**96.09**	88.16	**87.52 (8.54)**

Highest values are highlighted in boldface.

models. The verification accuracy for each subject and the averaged accuracy across these models are displayed in **Table 6**. From the average accuracy column, it is clear that only using the standard cross-entropy (in the case of $\alpha_{lsr} = 0$) for the model training is not an excellent choice. Proper selection of α_{lsr} can improve the CNN-based model's verification accuracy on the MI recognition. Second, we also observe that the model performance remains relatively stable across different values of α_{lsr} in a range of [0.25, 1]. This phenomenon suggests that different levels of smoothness on labels may have a similar effect on the model performance in the MI area. Finally, it can be seen that the model has the most remarkable improvements on subjects A02 and A05, who have low predicting accuracy, by using the smoothed cross-entropy compared to using the standard one. We further visualize the training and testing loss during the optimization of these two subjects in **Figure 3**. It is recognized that the models using standard cross-entropy suffer from an overfitting issue where the training loss decreases at the beginning and flattens gradually, but the testing loss decreases at the beginning while increases after several epochs. On the contrary, the models using smoothed cross-entropy have a good learning curve. Both training and testing errors decrease at the beginning and then flatten until the end of optimization. Together with the verification accuracy improvement, this finding suggests that the label smoothing technique can degrade the influence of the overfitting on the CNN model in the MI classification.

5.2.2. Center Loss

In the second experiment, we fix the α_{lsr} as 0.5 (performing the best in the first experiment) and vary α_{ct} in a range from 0 to 1 (inclusive) to train different models. The performances of these models are displayed in **Table 7**. When the α_{ct} is larger than zero, the center loss is activated, and the performance improves across almost all subjects and in the averaged level. This result suggests that the involvement of the center loss increases the discriminative power of the model. For a clear illustration and intuition, the principal component analysis (PCA) (Jolliffe and Cadima, 2016) is further employed to convert the high dimensional features of the second last layer in the model for subject A07 (randomly selected) into 2-D vectors. The distributions of these vectors are shown in **Figure 4**. It is clear that, without the center loss (**Figure 4A**), the deep features within each class are dispersive and have a larger intra-class variation. Alternatively, with the center loss in the joint supervision (**Figure 4B**), the features have both a compact intraclass distance and a clear interclass boundary. These results suggest that the center loss is beneficial to the discriminative feature learning for MI classification modeling.

TABLE 5 | Paired *t*-test (*p*-values) between our method and others.

Model	IV-2a	IV-2b
FBCSP	0.0002	0.0058
SMM	0.0005	0.0001
EEGNet	0.0443	0.0013
ConNet	0.0168	0.0228
DRDA	0.0120	0.0479

TABLE 6 | Classification accuracies (%) of models with different α_{lsr} values on the BCI competition IV-2a.

α_{lsr}	A01	A02	A03	A04	A05	A06	A07	A0	A09	Average
0	84.34	52.65	**93.04**	66.67	52.90	57.00	87.73	81.55	80.3	72.91
0.25	80.78	**64.66**	92.67	67.54	69.57	**60.00**	85.92	80.07	78.79	75.58
0.5	**84.34**	61.48	90.84	**72.37**	**73.19**	59.53	85.56	**82.66**	**81.44**	**76.82**
0.75	82.92	63.60	90.11	66.23	72.46	53.49	85.92	82.29	81.44	75.39
1	81.14	62.19	91.94	64.04	69.20	58.14	**89.53**	80.07	82.58	75.43

The α_{ct} is fixed as 0. The best accuracy for each subject is highlighted in boldface.

TABLE 7 | Classification accuracies (%) of models with different α_{ct} values on the BCI competition IV-2a.

α_{ct}	A01	A02	A03	A04	A05	A06	A07	A08	A09	Average
0	84.34	61.48	90.84	72.37	73.19	59.53	85.56	82.66	81.44	76.82
0.25	87.54	64.31	94.14	73.25	74.64	**63.26**	86.64	80.07	81.82	78.41
0.5	**89.32**	**66.78**	**94.14**	**74.56**	**76.45**	62.33	86.28	**85.61**	**85.23**	**80.08**
0.75	89.32	66.43	93.77	73.68	71.38	60.47	89.17	84.5 0	85.23	79.33
1	87.9	59.72	93.77	71.49	74.64	57.67	**89.89**	83.39	81.82	77.81

The α_{lsr} is fixed as 0.5. The best accuracy for each subject is highlighted in boldface.

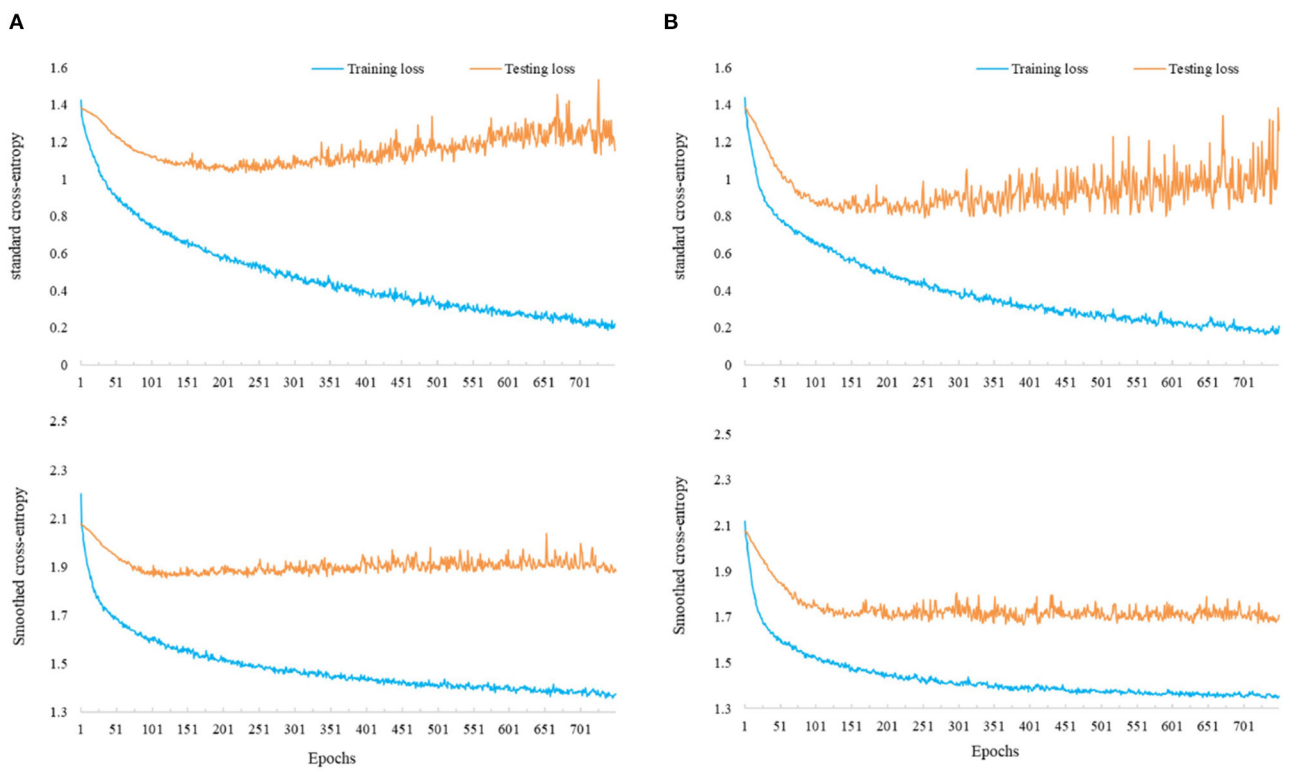


FIGURE 3 | Training and testing losses of the models using standard cross-entropy (top) and smoothed cross-entropy (bottom). **(A)** Learning curve of the model on A02. **(B)** Learning curve of the model on A05.

6. CONCLUSION

This paper proposes a new deep learning scheme for the CNN-based model in the MI classification. By jointly combining the smoothed cross-entropy with center loss, the robustness and discriminative power of the model can be highly enhanced for the classification. Extensive and systematic experiments are conducted to validate our strategy on two well-known

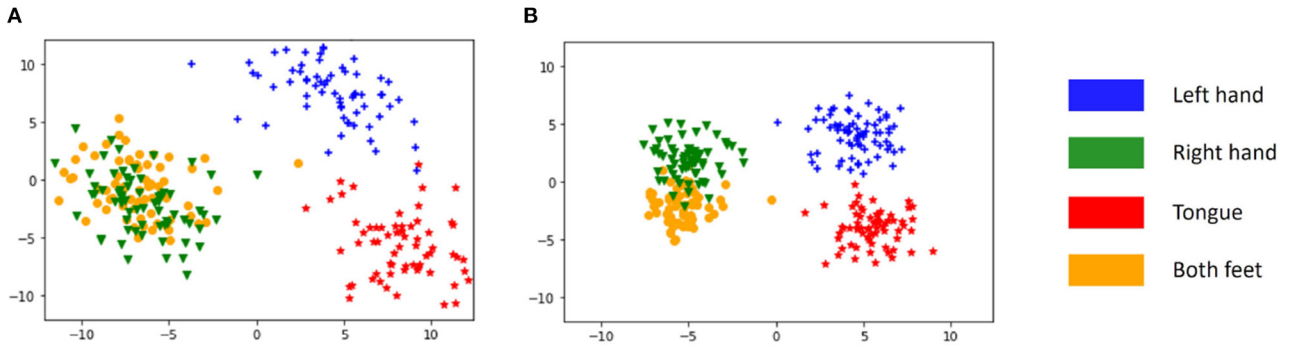


FIGURE 4 | Visualization of deep feature distributions of subject A07 on dataset IV-2a under two conditions. **(A)** Model with $\alpha_{lsr} = 0.5$ and $\alpha_{ct} = 0$. **(B)** Model with $\alpha_{lsr} = 0.5$ and $\alpha_{ct} = 0.5$.

benchmarks. Several suggestions have been made based on experimental findings. First, the label smoothing technique can degrade the overfitting issue caused by the scarcity and low SNR of the EEG data on the CNN model training. Second, the center loss along with the cross-entropy efficiently decreases the intra-class variance and thus increases the discriminative ability of the deep features by pulling them toward their corresponding latent class centers. It reduces the negative impact of the non-stationary characteristics of the EEG data on the MI classification task. Finally, the proposed loss offers a robust and discriminative training scheme for CNN-based modeling in the MI area. This phenomenon uncovers the fact that, in addition to sophisticated model structure development, implementing an efficient loss function for the learning guidance is also beneficial to model performance in MI recognition. This research has thus encouraged more attempts on the objective function innovation for the deep learning model in the MI field. It can be an interesting alternative for overcoming the bottleneck performance to the model architecture that has been heavily investigated.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: http://www.bbci.de/competition/iv/.

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.

AUTHOR CONTRIBUTIONS

XH conceptualized the study, performed the majority of the experiments and analyses, made the figures, and wrote the first draft of the manuscript. NZ and K-SC performed some experiments, updated the figures, performed the statistics, and edited the manuscript. All authors approved the submitted version.

FUNDING

The work in this paper was supported in part by the Hong Kong Innovation and Technology Fund (MRP/015/18), the Hong Kong Research Grants Council (PolyU 152006/19E), and National Nature Science Foundation of China (Nos. 61802036 and 11901063).

REFERENCES

Ahn, M., and Jun, S. C. (2015). Performance variation in motor imagery brain-computer interface: a brief review. *J. Neurosci. Methods* 243, 103–110. doi: 10.1016/j.jneumeth.2015.01.033

Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (Hong Kong), 2390–2397.

Bashivan, P., Rish, I., Yeasin, M., and Codella, N. (2015). Learning representations from EEG with deep recurrent-convolutional neural networks. *arXiv [Preprint]. arXiv:1511.06448*.

Bickel, S., Brückner, M., and Scheffer, T. (2009). Discriminative learning under covariate shift. *J. Mach. Learn. Res.* 10, 2137–2155. doi: 10.1145/1577069.1755858

Chorowski, J., and Jaitly, N. (2016). Towards better decoding and language model integration in sequence to sequence models. *arXiv [Preprint]. arXiv:1612.02695*. doi: 10.21437/Interspeech.2017-343

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv [Preprint]. arXiv:1511.07289*.

Goldenholz, D. M., Ahlfors, S. P., Hämäläinen, M. S., Sharon, D., Ishitobi, M., Vaina, L. M., et al. (2009). Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. *Hum. Brain Mapp.* 30, 1077–1086. doi: 10.1002/hbm.20571

Hersche, M., Rellstab, T., Schiavone, P. D., Cavigelli, L., Benini, L., and Rahimi, A. (2018). "Fast and accurate multiclass inference for mi-bcis using large multiscale temporal and spectral features," in *2018 26th European Signal Processing Conference (EUSIPCO)* (Rome), 1690–1694. doi: 10.23919/EUSIPCO.2018.8553378

Hertz, J., Krogh, A., and Palmer, R. G. (2018). *Introduction to the Theory of Neural Computation*. Redwood City: CRC Press. doi: 10.1201/9780429499661

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning* (Lille), 448–456.

Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A* 374:20150202. doi: 10.1098/rsta.2015.0202

Kaufmann, T., Herweg, A., and Kübler, A. (2014). Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials. *J. Neuroeng. Rehabil.* 11, 1–17. doi: 10.1186/1743-0003-11-7

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv [Preprint]. arXiv:1412.6980*.

Kumar, S., Sharma, A., Mamun, K., and Tsunoda, T. (2016). "A deep learning approach for motor imagery EEG signal classification," in *2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)* (Nadi), 34–39. doi: 10.1109/APWC-on-CSE.2016.017

Lal, T. N., Schroder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., et al. (2004). Support vector channel selection in BCI. *IEEE Trans. Biomed. Eng.* 51, 1003–1010. doi: 10.1109/TBME.2004.827827

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). Eegnet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* 15:056013. doi: 10.1088/1741-2552/aace8c

Lee, T.-W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719

Martinez-Leon, J.-A., Cano-Izquierdo, J.-M., and Ibarrola, J. (2015). Feature selection applying statistical and neurofuzzy methods to EEG-based BCI. *Comput. Intell. Neurosci.* 2015:781207. doi: 10.1155/2015/781207

Müller, R., Kornblith, S., and Hinton, G. (2019). When does label smoothing help? *arXiv [Preprint]. arXiv:1906.02629*.

Ni, T., Gu, X., and Jiang, Y. (2020). Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy. *J. Ambient Intell. Human. Comput.* 1–12. doi: 10.1007/s12652-020-02620-9

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134. doi: 10.1109/5.939829

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* 8, 441–446. doi: 10.1109/86.895946

Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. (2018). Brain-computer interface spellers: a review. *Brain Sci.* 8:57. doi: 10.3390/brainsci8040057

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.* 16:051001. doi: 10.1088/1741-2552/ab260c

Sakhavi, S., Guan, C., and Yan, S. (2018). Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 5619–5629. doi: 10.1109/TNNLS.2018.2789927

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* 38, 5391–5420. doi: 10.1002/hbm.23730

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas), 2818–2826. doi: 10.1109/CVPR.2016.308

Tabar, Y. R., and Halici, U. (2016). A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* 14:016003. doi: 10.1088/1741-2560/14/1/016003

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6:55. doi: 10.3389/fnins.2012.00055

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems* (Long Beach), 5998–6008.

Vidaurre, C., Klauer, C., Schauer, T., Ramos-Murguialday, A., and Müller, K.-R. (2016). Eeg-based bci for the linear control of an upper-limb neuroprosthesis. *Med. Eng. Phys.* 38, 1195–1204. doi: 10.1016/j.medengphy.2016.06.010

Wen, Y., Zhang, K., Li, Z., and Qiao, Y. (2016). "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision* (Amsterdam: Springer), 499–515. doi: 10.1007/978-3-319-46478-7_31

Wierzgała, P., Zapała, D., Wojcik, G. M., and Masiak, J. (2018). Most popular signal processing methods in motor-imagery BCI: a review and meta-analysis. *Front. Neuroinformatics* 12:78. doi: 10.3389/fninf.2018.00078

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Yu, H., Lu, H., Wang, S., Xia, K., Jiang, Y., and Qian, P. (2019). A general common spatial patterns for EEG analysis with applications to vigilance detection. *IEEE Access* 7, 111102–111114. doi: 10.1109/ACCESS.2019.2934519

Zhang, Y., Li, X., Zhu, J., Wu, C., and Wu, Q. (2019). Epileptic EEG signals recognition using a deep view-reduction tsk fuzzy system with high interpretability. *IEEE Access* 7, 137344–137354. doi: 10.1109/ACCESS.2019.2942641

Zhang, Y., Zhou, Z., Bai, H., Liu, W., and Wang, L. (2020). Seizure classification from EEG signals using an online selective transfer tsk fuzzy classifier with joint distribution adaption and manifold regularization. *Front. Neurosci.* 14:496. doi: 10.3389/fnins.2020.00496

Zhao, H., Zheng, Q., Ma, K., Li, H., and Zheng, Y. (2020). Deep representation-based domain adaptation for nonstationary EEG classification. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 535–545. doi: 10.1109/TNNLS.2020.3010780

Zheng, Q., Zhu, F., Qin, J., Chen, B., and Heng, P.-A. (2018). Sparse support matrix machine. *Pattern Recogn.* 76, 715–726. doi: 10.1016/j.patcog.2017.10.003

Zhou, H., and Li, L. (2014). Regularized matrix regression. *J. R. Stat. Soc. Ser. B* 76, 463–483. doi: 10.1111/rssb.12031

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/fnins.2021.773208

Kernelized *k*-Local Hyperplane Distance Nearest-Neighbor Model for Predicting Cerebrovascular Disease in Patients With End-Stage Renal Disease

Xiaobin Liu[1†], Xiran Zhang[1†], Yi Zhang[2†], Yijie Ding[3], Weiwei Shan[1], Yiqing Huang[1], Liang Wang[1] and Xiaoyi Guo[4*]*

[1] Department of Nephrology, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China, [2] NHC Key Laboratory of Nuclear Medicine, Jiangsu Key Laboratory of Molecular Nuclear Medicine, Jiangsu Institute of Nuclear Medicine, Wuxi, China, [3] Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou, China, [4] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Xiaoqing Gu, Changzhou University, China
Qiusheng Shen, Changshu Affiliated Hospital of Nanjing University of Traditional Chinese Medicine, China

***Correspondence:**
Liang Wang
wlwxsnk@163.com
Xiaoyi Guo
kerry.guoxiaoyi@163.com

[†] These authors share first authorship

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *09 September 2021*
Accepted: *04 October 2021*
Published: *25 October 2021*

Citation:
Liu X, Zhang X, Zhang Y, Ding Y, Shan W, Huang Y, Wang L and Guo X (2021) Kernelized k-Local Hyperplane Distance Nearest-Neighbor Model for Predicting Cerebrovascular Disease in Patients With End-Stage Renal Disease. Front. Neurosci. 15:773208. doi: 10.3389/fnins.2021.773208

Detecting and treating cerebrovascular diseases are essential for the survival of patients with chronic kidney disease (CKD). Machine learning algorithms can be used to effectively predict stroke risk in patients with end-stage renal disease (ESRD). An imbalance in the amount of collected data associated with different risk levels can influence the classification task. Therefore, we propose the use of a kernelized *k*-local hyperplane nearest-neighbor model (KHKNN) for the effective prediction of stroke risk in patients with ESRD. We compared our proposed method with other conventional machine learning methods, which revealed that our method could effectively perform the task of classifying stroke risk.

Keywords: cerebrovascular disease, end-stage renal disease, local hyperplane, klotho, FGF23

INTRODUCTION

Chronic kidney disease (CKD) has become a prominent disease affecting global health. According to existing research, the global incidence of CKD is approximately 8–16% and has been increasing yearly (Jha et al., 2013). Cerebrovascular diseases, such as stroke, represent major CKD complications that lead to neurological dysfunction and death, with negative impacts on prognosis in patients with CKD. Cerebral apoplexy, which is a primary cause of death among patients with CKD, refers to a series of adverse events, including cerebral ischemia, hypoxia, and cerebral dysfunction, caused by acute cerebral vascular rupture or acute cerebrovascular embolism (Kelly and Rothwell, 2020). A cohort study showed that CKD progression and a decline in the glomerular filtration rate increased the stroke risk among patients with CKD by nearly 40%, accompanied by a significant increase in the mortality rate (Toyoda and Ninomiya, 2014). Therefore, exploring the risk factors associated with stroke among the CKD population and identifying effective early interventions are necessary steps to reducing morbidity and mortality due to stroke.

Hypertension, diabetes, and dyslipidemia are traditional risk factors that contribute to the development of cerebrovascular diseases in patients with CKD. In addition, recent studies have revealed non-traditional risk factors, such as inflammation, oxidative stress, and CKD-mineral bone disease (CKD-MBD), that impact the occurrence and development of cerebrovascular diseases among patients with CKD. These non-traditional risk factors accelerate a series of

pathological processes, such as cerebrovascular endothelial injury and sclerosis, in patients with CKD, leading to cerebrovascular calcification, further changing hemodynamics, and ultimately causing cerebrovascular events (Allen and Bayraktutan, 2009).

Our previous work identified abnormal FGF23 and Klotho levels, inflammatory status, and malnutrition were the unconventional risk factors for vascular calcification and CKD-MBD in patients with end-stage renal failure (Maraj et al., 2018). Using machine learning methods to analyze the risk factors of CKD-MBD in patients with end-stage renal failure, we have found that elevated serum FGF23 levels in patients with ESRD is an independent risk factor for abdominal aortic calcification (Liu et al., 2021). Recent studies have also identified FGF23 as an independent risk factor for cerebrovascular diseases in both CKD and non-CKD populations (Wright et al., 2016). The CHADS2 (congestive heart failure, hypertension, age = 75 years, diabetes mellitus, stroke) and CHA2DS2-VASc (congestive heart failure, hypertension, age $\geq$ 75 years, diabetes mellitus, stroke or transient ischemic attack vascular disease, age 65 to 74 years, sex category) scores are currently well-recognized methods for predicting the risk of stroke in patients with CKD (Hsu et al., 2020). Therefore, in this study, based on previous research findings, we used machine learning algorithms to develop models that explore the scientificity and veracity of both traditional and non-traditional risk factors combined with the CHADS2 stroke scoring tool and an abdominal aortic calcification scoring method for the prediction of stroke risk in patients with ESRD, which could help clinicians identify cerebrovascular disease and provide early interventions by assessing various risk factors, potentially delaying the occurrence and development of stroke, reducing morbidity and mortality, and improving prognosis among patients with ESRD.

MATERIALS AND METHODS

Assessment of the CHADS2 and CHA2DS2-VASc Scores

We calculated the CHADS2 score based on the scoring system, as follows (de Bie et al., 2017): 1 point each was assigned for age $\geq$ 75 years, the presence of hypertension, diabetes mellitus, and congestive heart failure, and 2 points each were assigned for transient ischemic attack or a history of stroke. In addition, we calculated the CHA2DS2-VASc score based on the scoring system, as follows: 1 point each was assigned for congestive heart failure, hypertension, age between 65 and 74 years, diabetes mellitus, female sex, and vascular disease, whereas 2 points each were assigned for a history of stroke and age $\geq$ 75 years. CKD was defined as estimated glomerular filtration rate (eGFR) $<$ 60 mL/min/m^2 and classified as stages 3, 4, or 5 based on the eGFR level (30–59, 15–29, or $<$15 mL/min/1.73 m^2, respectively) combined with kidney damage lasting for longer than 3 months. Patients were categorized into two groups according to their CHADS2 and CHA2DS2-VASc risk scores: (1) low-risk group (0–1 score) for CHADS2 and CHA2DS2-VASc scores and (2) high-risk group ($\geq$2 scores) for CHADS2 and CHA2DS2-VASc scores.

Calculation of the Nutritional Indexes

The Geriatric Nutrition Risk Index (GNRI) = [14.89 $\times$ serum albumin (g/dl)] + [41.7 $\times$ (actual body weight/ideal body weight)] (Yamada et al., 2020). Serum levels of intact FGF23, klotho, fetuin-A, and interleukin-6 were determined using two-site enzyme-linked immunosorbent assays (Elabscience Biotech, Wuhan, China).

Abdominal Aortic Calcification Integration Method: Abdominal Aortic Calcification Score

All patients underwent lateral lumbar X-ray examinations within 1 week of biochemical blood examinations to assess abdominal aortic calcification corresponding to L1 to L4 (Asher et al., 2021). Each patient was scored based on the length of the calcified plaques identified on the anterior and posterior walls of the abdominal aorta, with each segment scored between 0 and 3 points, as follows: 0 points for no calcification; 1 point for calcification less than one-third of the arterial wall length; 2 points for calcification between one-third and two-thirds of the artery wall length; and 3 points if calcification covers more than two-thirds of the arterial wall length. Each lumbar segment is scored separately for both the posterior and anterior walls, resulting in a total score of 0–24 points. **Table 1** shows the demographic and clinical details of our data set.

The *k*-Local Hyperplane Distance Nearest-Neighbor Model

Vincent and Bengio developed an improved version of the *k*-nearest-neighbor algorithm (KNN), called the *k*-local hyperplane (LH) nearest-neighbor algorithm (HKNN) (Vincent and Bengio, 2002). The purpose of HKNN is to estimate the distance from the test sample in each class to its corresponding LH, which is built using the nearest k samples of the test sample. Suppose there are *C* classes; HKNN will obtain the predicted results of the test sample by calculating the minimum distance from *C* LHs. For the *c*-th class, the *c*-th LH is based on the nearest *k* neighbors of *x* in the training set, where *x* belongs to the *c*-th class. The *c*-th hyperplane is expressed as follows:

$$LH_k^c = \left\{ p^c | p^c = \overline{N}^c + \sum\nolimits_{i=1}^{k} \alpha_i^c V_i^c, \alpha_i^c \in R^k, i = 1, ..., k \right\} \quad (1)$$

where $\overline{N}^c = \frac{1}{k} \sum_{i=1}^{k} N_i^c$is the centroid of the k neighbors of x in class c; $N_i^c$is the ith neighbor of the test sample x in class c; and $V_i^c = N_i^c - \overline{N}^c$. The objective function of the test sample x to the cth LH is as follows:

$$\left(LH_k^c(x)\right)^2 = \left|\left| x - \overline{N}^c - \sum\nolimits_{i=1}^{k} \alpha_i^c V_i^c \right|\right|^2 + \lambda \sum\nolimits_{i=1}^{k} \left(\alpha_i^c\right)^2 \quad (2)$$

where λis the parameter of regular term. α^c can be calculated as:

$$\left(\left(V^c\right)^T V^c + \lambda I \right) \alpha^c = \left(V^c\right)^T \left(x - \overline{N}^c \right) \quad (3)$$

TABLE 1 | Data set information.

No.	Feature	Value	r*
1	Sex (male/female)	32/27	−0.0455
2	Age (years)	55.83 ± 15.60	0.4010
3	Smoking (yes/no)	1/58	−0.0847
4	BMI (kg/m^2)	23.56 ± 3.12	0.1639
5	DM (yes/no)	24/35	0.4847
6	CI (yes/no)	3/56	0.0025
7	CHD (yes/no)	5/54	0.0433
8	Systolic blood pressure (mmHg)	155.69 ± 23.63	−0.1150
9	Diastolic blood pressure (mmHg)	88.11 ± 13.71	−0.2043
10	Phosphate binder (yes/no)	36/23	−0.2141
11	Hemoglobin (g/L)	85.38 ± 18.12	−0.2584
12	C-reactive protein (mg/L)	11.74 ± 35.61	0.3016
13	Serum creatinine (μmol/L)	785.09 ± 368.62	−0.4252
14	Serum glucose (mmol/L)	5.48 ± 2.28	0.2608
15	Serum calcium (mmol/L)	2.08 ± 0.24	0.0520
16	Serum phosphorus (mmol/L)	1.81 ± 0.38	−0.0862
17	Total glyceride (mmol/L)	1.69 ± 1.08	−0.0542
18	Total cholesterol (mmol/L)	4.53 ± 1.42	−0.0466
19	Low density lipoprotein-C (mmol/L)	2.45 ± 0.96	−0.0252
20	High density lipoprotein-C (mmol/L)	0.97 ± 0.55	0.0866
21	HbA1c (%)	5.82 ± 1.03	0.2151
22	Serum albumin (g/L)	34.31 ± 6.61	−0.1308
23	25-OH Vitamin D3 (ng/ml)	7.86 ± 4.55	0.3850
24	iPTH (pg/ml)	274.50 ± 306.31	−0.0225
25	GNRI	96.06 ± 12.76	−0.0078
26	FGF23 (pg/ml)	32.21 ± 53.02	−0.0966
27	Klotho (ng/ml)	2.38 ± 2.33	0.0443
28	Interleukin-6 (pg/ml)	25.37 ± 53.69	0.2634
29	Fetuin-A (pg/ml)	3.0320e+05 ± 2.0606e+05	−0.0234
30	AACS	1.95 ± 1.55	0.2113
31	CHADS2 score	1.93 ± 1.11	0.4247
32	Group (low-risk group/high-risk group)	25/34	0.4619
33	CHA2DS2-VASc score	2.79 ± 1.50	0.5097

**Denotes each feature correlated with vascular calcification level using Pearson correlation coefficient (r).*
BMI, body mass index; DM, diabetes mellitus; CI, cerebral ischemia; CHD, coronary heart disease; HbA1c, glycated hemoglobin; iPTH, intact parathyroid hormone; GNRI, Geriatric Nutritional Risk Index; FGF23, fibroblast growth factor 23; AACS, Abdominal Aortic Calcification Score, CHADS2, congestive heart failure, hypertension, age = 75 years, diabetes mellitus, stroke; CHA2DS2-VASc, congestive heart failure, hypertension, age ≥ 75 years, diabetes mellitus, stroke or transient ischemic attack vascular disease, age 65–74 years, sex category.

where $\alpha^c = \left(\alpha_1^c, \alpha_2^c, ..., \alpha_k^c\right)^T$ and $V^c = \left(V_1^c, V_2^c, ..., V_k^c\right)_{d=k}$. The predictive result for test sample x is defined as:

$$class\ c = min\left|\left|x - \overline{N}^c - \sum_{i=1}^{k} \alpha_i^c V_i^c\right|\right|^2, c = 1, 2, ..., C \quad (4)$$

To further improve the performance of the model, we applied feature mapping and the kernel trick to HKNN to obtain a kernelized HKNN model (KNKNN). Let x map to f by $\phi : \chi \to F$ and set $\overline{x} = x - \overline{N}^c$. Eq. 2 can then be rewritten, as follows:

$$\left(LH_k^c(x)\right)^2 = \left|\left|\phi(\overline{x}) - \sum_{i=1}^{k} \alpha_i^c \phi\left(V_i^c\right)\right|\right|^2 + \lambda \sum_{i=1}^{k} \left(\alpha_i^c\right)^2 \quad (5)$$

We obtained the differential of Eq. 5 as follows:

$$\left(\left(\phi\left(V^c\right)\right)^T \phi\left(V^c\right) + \lambda I\right) \alpha^c = \left(\phi\left(V^c\right)\right)^T \phi(\overline{x}) \quad (6a)$$

$$\alpha^c = \left(\left(\phi\left(V^c\right)\right)^T \phi\left(V^c\right) + \lambda I\right)^{-1} \left(\phi\left(V^c\right)\right)^T \phi(\overline{x}) \quad (6b)$$

$$\alpha^c = \left(K\left(V^c, V^c\right) + \lambda I\right)^{-1} K\left(V^c, \overline{x}\right) \quad (6c)$$

where $K\left(V^c, V^c\right) \in R^{k=k}$ is a Gram matrix calculated by the radial basis function (RBF), and $K\left(V^c, \overline{x}\right) \in R^{k=1}$ is a vector. The RBF is defined as:

$$K\left(x_i, x_j\right) = exp\left(-\gamma \parallel x_i - x_j \parallel^2\right) \quad (7)$$

where γ is the Gaussian kernel bandwidth.

To avoid overfitting of the model, KHKNN employed two strategies: (1) For test sample, KHKNN separately constructs a local hyperplane for each category by linear representation of neighborhood samples. It can alleviate the parameter (number) sensitivity of neighbors. (2) When constructing the hyperplane, we added the regular term (L2) of the coefficient. The schematic diagram of KNKNN is shown in **Figure 1**.

RESULTS

Measurements

In this study, accuracy (ACC), sensitivity (SN), specificity (SP), positive predictive value (PE), negative predictive value (NPV), a weighted average of the PE and sensitivity (F_{score}), and Matthews correlation coefficient (MCC) were calculated as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (8a)$$

$$SN = \frac{TP}{TP + FN} \quad (8b)$$

$$Spec = \frac{TN}{TN + FP} \quad (8c)$$

$$PE = \frac{TP}{TP + FP} \quad (8d)$$

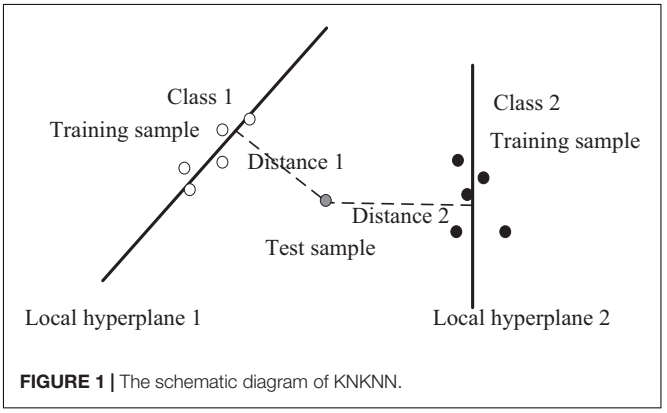


FIGURE 1 | The schematic diagram of KNKNN.

$$NPV = \frac{TN}{TN + FN} \tag{8e}$$

$$F_{score} = 2 \times \frac{SN \times PE}{SN + PE} \tag{8f}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN) \times (TN + FP) \times (TP + FP) \times (TN + FN)}} \tag{8g}$$

where *TP*, *TN*, *FN*, and *FP* are the number of true positive, true negative, false negative, and false positive results, respectively. The high-risk patients are positive samples, whereas the low-risk patients are negative samples.

The Optimal Parameters

In our model, three parameters (k, γ, and λ) must be selected. To make the model robust, we set λ as 1. The optimal parameters of k and γ were selected using a grid search. The value of k ranges from 2 to 8 (maximum number of negative samples), with a step of 1. The value of γ ranges from 2^{-5} to 2^{5} with a step of 2^{1}. The results are shown in **Figure 2**, which shows the predictive performance of the model using different parameters. When k and $\gamma$are 4 and 2^{-3}, respectively, the best MCC (0.5393) value was obtained.

Comparison of Performance Between Our Method and Other Existing Methods

We compared the performance of several traditional classifiers with our method using the obtained dataset. The results are shown in **Table 2**. The compared classifiers included artificial neural network with backpropagation (ANN-BP), support vector machines (SVM), Takagi-Sugeno-Kang fuzzy system (TSK-FS), KNN, HKNN, and KHKNN. For small samples, KNN (ACC: 86.67%), HKNN (ACC: 85.00%), and TSK-FS (ACC: 78.33%) achieved good results. Our method (KHKNN) achieved the best MCC (0.5393) and ACC (89.67%) on small dataset. KHKNN was found to have an SN of 94.00%, with Spec reaching 60.00%. Our method achieved a relatively balanced performance for the recognition of both positive and negative samples.

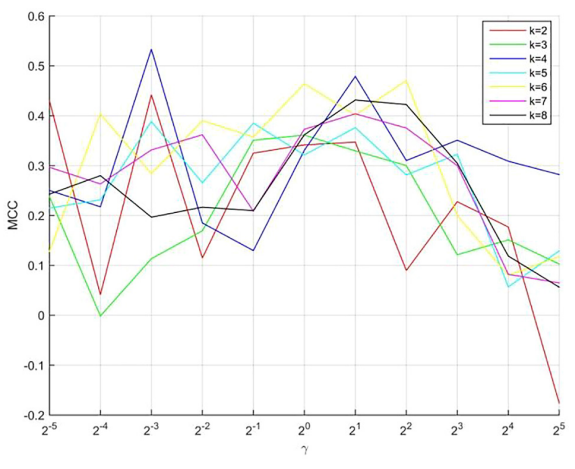


FIGURE 2 | The predictive performance of the model using different parameters.

DISCUSSION

Stroke is one of the most serious complications among patients with CKD, leading to brain dysfunction and even death. Over the past 10 years, scholars have conducted a large number of mechanistic studies and epidemiological investigations exploring the kidney–brain interaction. The results of these studies have indicated that the kidney and brain have similar anatomical and functional characteristics. For example, both organs feature an arterial system that automatically adjusts perfusion pressure to ensure a continuous and relatively stable blood flow. In patients with CKD, cerebrovascular sclerosis occurs due to calcifications that form in the arterial system, disrupting the autoregulation function and allowing cerebrovascular events to occur (Lau et al., 2017).

The risk of stroke in patients with CKD is much higher than that in patients without CKD (Chen et al., 2012), and the stroke risk increases further as renal functional defects progress to ESRD. In recent years, in addition to traditional risk factors, such as hypertension, diabetes, and dyslipidemia, the influence of non-traditional risk factors on the occurrence of cerebrovascular calcification in patients with CKD has gained increasing attention, including inflammation, malnutrition, and the FGF23/klotho axis. The results of previous studies performed at our center have indicated that abnormal FGF23, klotho, and fetuin-A levels and malnutrition represent risk factors for abdominal aortic calcification in patients with ESRD (Maraj et al., 2018). FGF23 has been to play an important role in phosphate regulation. Klotho is the receptor protein for FGF23, which participates in regulating bone, calcium, and phosphorus metabolism; protecting the integrity of blood vessels; and inhibiting vascular calcification through the formation of FGF23-klotho complexes. The FGF23/klotho axis is a key participant in CKD-MBD and is closely related to vascular calcification and cerebrovascular diseases (Moldovan et al., 2014). Relevant studies have shown that an elevated FGF23 level is a risk factor for ischemia and hemorrhagic stroke in patients with CKD (Wright et al., 2014).

TABLE 2 | Comparison of performance between our method and other existing methods using the PDB1075 data set (Jackknife test evaluation).

Methods	MCC	ACC (%)	SN (%)	Spec (%)	PE (%)	NPV (%)	F_{score}
ANN-BP	0.3521	77.58	82.18	**60.00**	92.32	40.00	0.8541
SVM	NaN	86.52	100	0	86.52	NaN	0.9272
TSK-FS	0.3803	78.33	82.91	**60.00**	92.00	43.33	0.8612
KNN	0.3697	86.67	**94.18**	40.00	90.73	50.00	0.9237
HKNN	0.4865	85.00	90.55	50.00	91.92	**66.67**	0.9094
KHKNN	**0.5393**	**89.67**	94.00	**60.00**	**94.52**	63.33	**0.9407**

MCC, Matthews correlation coefficient; ACC, accuracy; SN, sensitivity; Spec, specificity; PE, positive predictive value; NPV, negative predictive value; F_{score}, weighted average of the PE and sensitivity; ANN-BP, artificial neural network with backpropagation; SVM, support vector machines; TSK-FS, Takagi-Sugeno-Kang fuzzy system; KNN, k-nearest-neighbor; HKNN, k-local hyperplane distance nearest-neighbor algorithm; KHKNN, kernelized HKNN; NaN, not a number. The best results in each column are in bold faces.

The CHADS2 and CHA2DS2-VASc scores are considered to be effective evaluation tools for predicting the risk of ischemic stroke in patients with CKD. Among patients with a high risk of ischemic cerebrovascular accidents, the administration of secondary prevention agents, such as anticoagulation and antithrombosis factors, when indicated by the cerebral ischemia score warning system, has been shown to greatly reduce the incidence of stroke and improve prognosis (Toyoda et al., 2014). The results of this study revealed a correlation between the CHADS2 and CHA2DS2-VASc scores in patients with ESRD and vascular calcification, which indirectly suggests the existence of an important relationship between vascular calcification and ischemic stroke. Moreover, when we combined the ischemic stroke scoring tool with traditional stroke risk factors, such as vascular calcification, to predict the risk of CKD stroke, the results were more reasonable, with a stronger scientific basis, than the use of the stroke scoring tool alone to predict risk.

The sample size is very small, so a simple machine learning model is preferred to solve the classification problem. Among them, KNN and SVM are suitable methods. The KNN algorithm is very sensitive to the number of neighboring samples. In the original feature space, the model cannot achieve satisfactory results of prediction. Therefore, we proposed KHKNN on the basis of KNN and HKNN to solve the above two problems. In the results section, KHKNN has obtained good prediction results. KHKNN separately constructs a local hyperplane for each category of test sample. The prediction result is determined by evaluating the distance (minimum) from the test sample to the hyperplane of each category. Therefore, it can alleviate the parameter sensitivity of KNN and avoid overfitting. The prediction result is determined by evaluating the distance (minimum) from the test sample to the hyperplane of each category.

In our study, KHKNN was employed to predict the risk of cerebrovascular disease among patients with ESRD. KHKNN estimates the distance from the test sample in each class to its corresponding LH in a high-dimensional feature space. Unlike KNN, KHKNN is not as sensitive to the parameter k, and its prediction performance is better than that of the ANN-BP and SVM models for small data sets. Compared with other models (ANN-BP, SVM, TSK-FS, KNN, and HKNN), our model achieved the best MCC (0.5393) and ACC (89.67%) values, showing that our method has good robustness and may be useful for determining clinical risk in the future.

CONCLUSION

We proposed a KHKNN method to filter noise samples, improve the generalization ability of the model, and obtain good results. Although our method achieves a relatively balanced performance for the recognition of positive and negative samples, the following disadvantages must be acknowledged. (1) The sample size must be further increased to minimize prediction bias. (2) No detailed analysis was performed to examine the contribution of various patient factors. (3) Although the kernel function was used to map the original space to further improve the performance, the interpretability of the model was affected. Fuzzy systems will be introduced in the future to improve interpretability. At present, artificial intelligence technology has been used for large-scale medical information processing (Jian et al., 2019; Guo et al., 2021; Jiang et al., 2021a,b; Zhang et al., 2021a,b) and bioinformatics (Qian et al., 2021; Zou et al., 2021) on a large scale, with good performance. (4) The k-dimensional tree is employed to speed up the search speed of the nearest neighbor samples. In addition, parallel computing technology also can increase the speed of searching. In the future, we will use artificial intelligence methods to solve additional clinical problems.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Ethics Committee (Wuxi People's Hospital Ethics Committee, No. KS2019041). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XL: methodology, data curation, and writing – original draft preparation. XZ: methodology and writing – original draft preparation. WS and XG: methodology and data curation. YD: methodology and software. LW: methodology, supervision, and writing – reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by a grant from the Top Talent Support Program for young and middle-aged people of Wuxi Health Committee [HB2020008]; the Scientific Research Project of Wuxi Health Committee [MS201927; Z201914]; the Scientific Research Project of Jiangsu Provincial Health Commission [LGY201801]; Jiangsu Province "333" project [BRA2020142]; Medical and Public Health Project of Wuxi Sci-Tech Development Fund [WX18 II AN047]; and Maternal and Child Health Research Project of Jiangsu Province [F202033].

ACKNOWLEDGMENTS

We thank the Department of Nephrology of Wuxi People's Hospital for collecting data in this study.

REFERENCES

Allen, C. L., and Bayraktutan, U. (2009). Oxidative stress and its role in the pathogenesis of ischaemic stroke. *Int. J. Stroke.* 4, 461–470. doi: 10.1111/j.1747-4949.2009.00387.x

Asher, E., Abu-Much, A., Bragazzi, N. L., Younis, A., Younis, A., Masalha, E., et al. (2021). CHADS2 and CHA2DS2-VASc scores as predictors of platelet reactivity in acute coronary syndrome. *J. Cardiol.* 77, 375–379. doi: 10.1016/j.jjcc.2020.09.010

Chen, Y. C., Su, Y. C., Lee, C. C., Huang, Y. S., and Hwang, S. J. (2012). Chronic kidney disease itself is a causal risk factor for stroke beyond traditional cardiovascular risk factors: a Nationwide Cohort Study in Taiwan. *PLoS One* 7:e36332. doi: 10.1371/journal.pone.0036332

de Bie, M. K., Buiten, M. S., Rotmans, J. I., Hogenbirk, M., Schalij, M. J., Rabelink, T. J., et al. (2017). Abdominal aortic calcification on a plain X-Ray and the relation with significant coronary artery disease in asymptomatic chronic dialysis patients. *BMC Nephrol.* 18:82. doi: 10.1186/s12882-017-0480-2

Guo, X., Zhou, W., Shi, B., Wang, X., Du, A., Ding, Y., et al. (2021). An efficient multiple kernel support vector regression model for assessing dry weight of hemodialysis patients. *Curr. Bioinform.* 16, 284–293. doi: 10.2174/1574893615999200614172536

Hsu, P. C., Lee, W. H., Chen, S. C., Tsai, Y. C., Chen, Y. C., Chu, C. Y., et al. (2020). Using CHADS2 and CHA2DS2-VASc Scores for mortality prediction in patients with chronic kidney disease. *Sci. Rep.* 10:18942. doi: 10.1038/s41598-020-76098-y

Jha, V., Garcia-Garcia, G., Iseki, K., Naicker, S., Plattner, B., Saran, R., et al. (2013). Chronic kidney disease: global dimension and perspectives. *Lancet* 382, 260–272. doi: 10.1016/S0140-6736(13)60687-X

Jian, Y., Zhao, K., Xia, K., Xue, J., Xhou, L., Ding, Y., et al. (2019). A novel distributed multitask fuzzy clustering algorithm for automatic mr brain image segmentation. *J. Med. Syst.* 43:118. doi: 10.1007/s10916-019-1245-1

Jiang, Y., Gu, X., Wu, D., Hang, W., Xue, J., Qiu, S., et al. (2021a). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. *IEEE ACM Trans. Comput. Biol. Bioinform.* 18, 40–52. doi: 10.1109/TCBB.2019.2963873

Jiang, Y., Zhang, Y., Lin, C., Wu, D., and Lin, C.-T. (2021b). EEG-Based Driver Drowsiness Estimation Using an Online Multi-View and Transfer TSK Fuzzy System. *IEEE Trans. Intell. Transp. Syst.* 22, 1752–1764. doi: 10.1109/TITS.2020.2973673

Kelly, D., and Rothwell, P. M. (2020). Disentangling the Multiple Links Between Renal Dysfunction and Cerebrovascular Disease. *J. Neurol. Neurosurg. Psychiatry* 91, 88–97. doi: 10.1136/jnnp-2019-320526

Lau, W. L., Huisa, B. N., and Fisher, M. (2017). The cerebrovascular-chronic kidney disease connection: perspectives and mechanisms. *Transl. Stroke Res.* 8, 67–76. doi: 10.1007/s12975-016-0499-x

Liu, X. B., Zhang, X. R., Guo, X. Y., Ding, Y., Shan, W., Wang, L., et al. (2021). A self-representation-based fuzzy svm model for predicting vascular calcification of hemodialysis patients. *Comput. Math Methods Med.* 2021:2464821. doi: 10.1155/2021/2464821

Maraj, M., Kuśnierz-Cabala, B., Dumnicka, P., Gala-Błądzińska, A., Gawlik, K., Pawlica-Gosiewska, D., et al. (2018). Malnutrition, inflammation, atherosclerosis syndrome (MIA) and diet recommendations among end-stage renal disease patients treated with maintenance hemodialysis. *Nutrients* 10:69. doi: 10.3390/nu10010069

Moldovan, D., Moldovan, I., Rusu, C., Kacso, I., Patiu, I. M., and Gherman-Caprioara, M. (2014). FGF-23, vascular calcification, and cardiovascular diseases in chronic hemodialysis patients. *Int. Urol. Nephrol.* 46, 121–128. doi: 10.1007/s11255-013-0422-2

Qian, Y., Meng, H., Lu, W., Liao, Z., Ding, Y., and Wu, H. (2021). Identification of DNA-binding proteins via hypergraph based laplacian support vector machine. *Curr. Bioinform.* doi: 10.2174/1574893616666210806091922 [Epub ahead of print].

Toyoda, K., and Ninomiya, T. (2014). Stroke and cerebrovascular diseases in patients with chronic kidney disease. *Lancet Neurol.* 13, 823–833. doi: 10.1016/S1474-4422(14)70026-2

Toyoda, K., Yasaka, M., Uchiyama, S., Iwade, K., Koretsune, Y., Nagata, K., et al. (2014). CHADS2 and CHA2DS2-VASc scores as bleeding risk indices for patients with atrial fibrillation: the bleeding with antithrombotic therapy study. *Hypertens. Res.* 37, 463–466. doi: 10.1038/hr.2013.150

Vincent, P., and Bengio, Y. (2002). K-local hyperplane and convex distance nearest neighbor algorithms. *Adv. Neural Inf. Process. Syst.* 14, 985–992.

Wright, C. B., Dong, C., Stark, M., Silverberg, S., Rundek, T., Elkind, M. S., et al. (2014). Plasma FGF23 and the risk of stroke: the northern manhattan s tudy (NOMAS). *Neurology* 82, 1700–1706. doi: 10.1212/WNL.0000000000000410

Wright, C. B., Shah, N. H., Mendez, A. J., DeRosa, J. T., Yoshita, M., Elkind, M. S., et al. (2016). Fibroblast growth factor 23 is associated with subclinical cerebrovascular damage: the northern manhattan study. *Stroke* 47, 923–928. doi: 10.1161/STROKEAHA.115.012379

Yamada, S., Yamamoto, S., Fukuma, S., Nakano, T., Tsuruya, K., and Inaba, M. (2020). Geriatric nutritional risk index (GNRI) and creatinine index equally predict the risk of mortality in hemodialysis patients: J-DOPPS. *Sci. Rep.* 10:5756. doi: 10.1038/s41598-020-62720-6

Zhang, Y. P., Jian, Y., Qi, L., Bhuiyan, M. Z. A., and Quian, P. (2021a). Epilepsy diagnosis using multi-view & multi-medoid entropy-based clustering with privacy protection. *ACM Trans. Internet Technol.* 21, 1–20. doi: 10.1145/3404893

Zhang, Y. P., Wang, S. H., Xia, K. J., Jiang, Y. Z., and Qian, Q. J. (2021b). alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inf. Fusion.* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002

Zou, Y., Wu, H., Guo, X., Peng, L., Ding, Y., Tang, J., et al. (2021). MK-FSVM-SVDD: a multiple kernel-based fuzzy SVM model for predicting DNA-binding proteins via support vector data description. *Curr. Bioinform.* 16, 274–283. doi: 10.2174/1574893615999200607173829

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

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

Copyright © 2021 Liu, Zhang, Zhang, Ding, Shan, Huang, 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 in Neuroscience

ORIGINAL RESEARCH
published: 28 October 2021
doi: 10.3389/fnins.2021.780147

A Novel Active Rehabilitation Model for Stroke Patients Using Electroencephalography Signals and Deep Learning Technology

*Ming Gao[1] and Jie Mao[2]**

[1] College of Sports Science and Technology, Wuhan Sports University, Wuhan, China, [2] College of Sports Engineering and Information Technology, Wuhan Sports University, Wuhan, China

The main clinical manifestations of stroke are motor, language, sensory, and mental disorders. After treatment, in addition to being conscious, other symptoms will still remain in varying degrees. This is the sequelae of stroke, including numbness, facial paralysis, central paralysis, and central paralysis. If the sequelae of stroke are not treated effectively, they can easily develop into permanent sequelae. Most of the affected people have sequelae, and most of them have symptoms of upper limb paralysis. Therefore, it is of great significance to study how to carry out effective rehabilitation training for stroke patients to reduce the disease and even restore their motor function. Based on this background, this research aims to use deep learning technology to design a stroke rehabilitation model based on electroencephalography (EEG) signals. First, the patient's EEG signal will be preprocessed. Then, an improved deep neural network model (IDNN) is used to get the EEG classification results. The traditional DNN model construction process is simple and suitable for scenarios where there is no special requirement for the data format, but the generalization of a single DNN model is usually poor. Large margin support vector machine (LM_SVM) is an extension method of support vector machine (SVM), suitable for any occasion. By optimizing the edge distribution, better generalization performance can be obtained. Taking into account the advantages of DNN and LM_SVM and the high aliasing characteristics of stroke data, an improved DNN model is proposed. Finally, based on the EEG recognition result of the model, the rehabilitation equipment is controlled to assist the patient in rehabilitation treatment. The experimental results verify the superiority of the EEG classification model used, and further prove that this research has good practical value.

Keywords: stroke, rehabilitation model, deep learning technology, deep neural network model, EEG recognition

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Yanhui Zhang,
Hebei University of Chinese Medicine, China
Xiaoqing Gu,
Changzhou University, China

****Correspondence:***
Jie Mao
maojie@whsu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *20 September 2021*
Accepted: *11 October 2021*
Published: *28 October 2021*

Citation:
Gao M and Mao J (2021) A Novel Active Rehabilitation Model for Stroke Patients Using Electroencephalography Signals and Deep Learning Technology. Front. Neurosci. 15:780147. doi: 10.3389/fnins.2021.780147

INTRODUCTION

Stroke is a disease that causes brain dysfunction due to blockage or rupture of blood vessels in the brain. According to the cerebral blood circulation disorder, it is divided into two types: hemorrhagic type and ischemic type. Most stroke patients are ischemic, that is, the softening and necrosis of local brain tissue due to blood circulation, ischemia, and hypoxia. There are more ischemic strokes,

accounting for about 80% of all strokes (Liu and Chen, 2014). According to the statistics of the World Health Organization, stroke has become one of the main diseases of human death and disability worldwide (Sheorajpanday et al., 2011; Walter et al., 2016). The age of patients tends to be younger. Due to the different extent of damage to the affected brain regions and tissues, the symptoms of stroke also vary. Symptoms include paralysis of limbs, impaired language and communication skills, loss of comprehension, and loss of emotional thinking. Among them, post-stroke cognitive dysfunction (PSCI) is a common complication (Pendlebury et al., 2015). According to different races, regions and diagnostic criteria, the incidence of PSCI in Europe and America is 20∼80% (Douiri et al., 2013; Sun et al., 2014).

In recent years, with the emergence of new treatment methods such as computer-assisted cognitive training, new drug development, and occupational therapy, it has brought hope to stroke rehabilitation. These treatment methods have improved the quality of rehabilitation, shortened the recovery period, and reduced the cost of rehabilitation (Lin et al., 2014). Topping (2002) developed a robot that can assist children with cerebral palsy to eat. This is an assistive robot, and its upgraded version can also provide services such as cleaning and brushing. Krebs et al. (1998) has developed a robot that can assist rehabilitation training based on brain nerves. This robot can perform upper limb rehabilitation training for stroke patients. Compared with traditional therapy, it has achieved good results. Leigh et al. (2006) collects EEG by implanting a chip in the sample brain. The samples can realize a series of behaviors in daily life through the brain-computer interface, such as checking emails, turning on the TV, controlling the damaged upper limbs to grasp and so on. This is of great significance to the research direction of using ideas to control external devices, and it also provides new ideas for rehabilitation training based on brain-computer interface technology. Gomez-Rodriguez et al. (2011) developed a rehabilitation system based on motor imagination. The system first records the EEG signals when the sample imagines the arm movement, and then recognizes the movement tendency of the sample online, and drives the robotic arm to move based on the recognition result. Based on the existing research, it can be seen that the rehabilitation of stroke mainly relies on medical rehabilitation equipment. In order to assist patients in active rehabilitation training, some active rehabilitation methods have been proposed. The core idea of this method is to control the rehabilitation equipment based on the patient's EEG signal recognition result, and then assist the patient to train their consciousness and limbs synchronously. Gert et al. (2003) designed a MI-EEG-based rehabilitation system for the recovery of patients' arm motor function.

In the rehabilitation research of stroke samples, the most important thing is to classify and recognize EEG signals. The classification accuracy of EEG directly determines the rehabilitation effect. EEG recognition research is mainly divided into two categories, which are based on machine learning (Jiang et al., 2019, 2021) and deep learning (Elman, 1990; Dahl et al., 2012). The core idea of EEG recognition based on machine learning is to preprocess the collected EEG signals, then feature extraction, and finally use the trained classifier to recognize feature data to obtain the classification results. The commonly used feature extraction methods in this process are time domain analysis method (Fingelkurts et al., 2007), frequency domain analysis method (Anderson et al., 1995), time-frequency domain analysis method (Zandi et al., 2010), and so on. Commonly used classifiers are SVM (Garrett et al., 2003), Linear Discriminant Analysis (LDA; Cortes and Vapnik, 1995), BP neural network (Williams et al., 1986). The core idea of EEG recognition based on deep learning is to use various deep learning algorithms to directly classify EEG signals. Commonly used deep learning models are Deep Neural Networks (DNN; Ghonchi et al., 2020), Deep Belief Networks (DBN; Leigh et al., 2006), Auto Encoder (AE; Li et al., 2021), Spare Coding (SC; Tokuhiro, 2000), Convolutional Neural Networks (CNN; Deng et al., 2009) and Recurrent Neural Network (RNN; Petrosian et al., 2000), Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber, 1997).

Deep neural network uses the output of the previous layer as the input of the next layer to obtain the hierarchical characteristics of the input data. Finally, by adjusting the parameters in the neural network layer, the difference between the output and the input is minimized. The DNN model construction process is relatively simple and is suitable for scenarios where there is no special requirement for the data format, but usually a single DNN model has poor generalization. LM_SVM is suitable for any occasion based on SVM. It obtains better generalization performance by optimizing the edge distribution. The minimum margin of LM_SVM is usually smaller than that of SVM, and the margin distribution is better than SVM. In most cases, LM_SVM usually produces a larger margin than SVM. By maximizing the edge mean and variance at the same time, LM_SVM can effectively process data with high aliasing. Considering the advantages of DNN and LM_SVM and the high aliasing characteristics of stroke data, an improved DNN (IDNN) is proposed. The EEG signal recognition result based on IDNN can effectively control the rehabilitation equipment. The core work of this research is:

(1) An improved DNN model is proposed: This model uses DNN as the kernel function of LM_SVM to perform kernel mapping on the input data. Then the output of DNN is fed back to LM_SVM, and LM_SVM outputs the final classification result of the model. IDNN makes up for the shortcomings of using DNN alone, so that the constructed model has better generalization. As the core part of LM_SVM, the kernel function directly affects the classification performance of the model. In this study, DNN is used as a kernel function to solve the problem of kernel selection.
(2) EEG recognition is based on the IDNN model: Compared with the classification results of other models, the model used has a higher recognition rate. Applying EEG recognition results to rehabilitation training equipment for stroke, patients can control the rehabilitation equipment to complete corresponding actions more quickly and accurately. This rehabilitation training model can not only

train the thinking but also assist in the completion of the movement of the limbs, and has good practical value.

RELATED WORK

Stroke Rehabilitation Treatment Process

At present, China has established a three-level rehabilitation system for rehabilitation of stroke samples. The difference between each level of rehabilitation lies mainly in the distinction of rehabilitation sites. "First-level rehabilitation" refers to early rehabilitation treatment and conventional treatment received in the emergency room or neurology ward. The training of the ability to sit down and the recovery of eating ability lays the foundation for the sample to leave the stroke ward for upgraded rehabilitation. "Secondary rehabilitation" means that the sample enters the rehabilitation ward or rehabilitation center. At this stage, clinical treatment is used to assist rehabilitation treatment, mainly for rehabilitation treatment. The main purpose of this stage is to improve the sample's physical ability and daily living ability. After training at this stage, most samples can take care of themselves and return to their families. "Tertiary rehabilitation" refers to the treatment in which the sample leaves the hospital and returns to the community or home. Through further daily life and sports training to consolidate recovery.

In "first-level rehabilitation," most of the samples are in a coma or bedridden state, and only basic early recovery is possible. Rehabilitation treatment is less involved. Rehabilitation treatment mainly includes rehabilitation care, recovery of consciousness, management of swallowing function, and passive activities with the help of others. In the "three-level rehabilitation," the place of rehabilitation is changed from a hospital to a home or a community, and the practitioners of auxiliary treatment are physicians. This study is mainly used for the "secondary rehabilitation" treatment stage where rehabilitation treatment is more involved and the location is the rehabilitation center.

The secondary rehabilitation process for stroke is shown in **Figure 1**. After the sample is transferred to the rehabilitation center or the rehabilitation department of the hospital, the rehabilitation physician collects the relevant medical history of the patient and conducts specialized examination and functional screening of the patient. Mainly check the degree of obstacles in motor function, sensory function, cognitive communication, and self-care ability in life. Determine the composition of the rehabilitation team based on the dysfunction. Further inspection of the samples to determine the degree of disease and staging results. Then the patients are evaluated for rehabilitation, and rehabilitation treatment plans are developed and treatments are carried out. The training content at this stage is mainly for the training of daily life functions such as sitting, standing and walking, eating and dressing, and going up and down stairs. After a period of training, evaluate the rehabilitation effect of the sample. If the sample's dysfunction has improved, it can enter the community for rehabilitation. If the patient does not meet the return to community living standards, it is recommended to continue to be hospitalized for rehabilitation.

Deep Neural Network

This section introduces the related principles of the neural network model involved in the stroke assisted rehabilitation model mentioned. DNN is a feed-forward artificial neural network that contains multiple hidden layers between the input and output layers. Compared with ordinary neural networks, deep learning networks require unsupervised learning before supervised learning, and then use the learned weights as initial values for supervised learning for training. This supervised learning helps reduce the risk of overfitting. The DNN model structure is shown in **Figure 2**.

The DNN shown in **Figure 2** uses a multi-layer neural network to simulate the human brain to learn features automatically. Extract features from the original data to analyze the data. The basic idea of deep learning is to use a multi-layer neural network, and the output of the previous layer is used as the input of the next layer. Based on this method, the hierarchical representation of the input data is realized, and the hierarchical characteristics of the input data are obtained. Finally, by adjusting the parameters in the neural network layer, the difference between the output and the input is as small as possible. The implementation steps of DNN are described as follows:

(1) Unsupervised learning is used to pre-train each layer of the network;
(2) Based on the unsupervised learning training of each layer of neural network, the training results obtained will be used as the input of the next layer;
(3) A top-down supervision algorithm is used to adjust the parameters of the neural network layer.

Since deep learning was proposed, it has attracted the attention of many researchers. After more than 10 years of development, deep learning technology has been widely used in many fields of data mining. Many impressive results have been achieved in pattern recognition, intelligent robots, automatic control, predictive estimation, biology, medicine, and economics.

Video behavior recognition currently exists in algorithms, both manual feature-based, and deep learning-based methods, to extract spatial and temporal information from the video, spatial information contains information about the appearance of the video, and temporal information contains information about the temporal continuity between frames of the video. The spatial information is usually extracted using a convolutional neural network, and the temporal information is usually extracted using a cyclic neural network, which uses different filters to get different picture features. Different convolutional kernels correspond to different picture effects.

LM_SVM

LM_SVM is an extension method of SVM, which is suitable for any occasion that supports SVM. LM_SVM obtains better generalization performance by optimizing the edge distribution. The minimum margin of LM_SVM is usually larger than that of SVM, and the margin distribution is better than SVM. Therefore, compared to SVM, LM_SVM has more advantages in processing data with high aliasing. The SVM algorithm based on edge theory

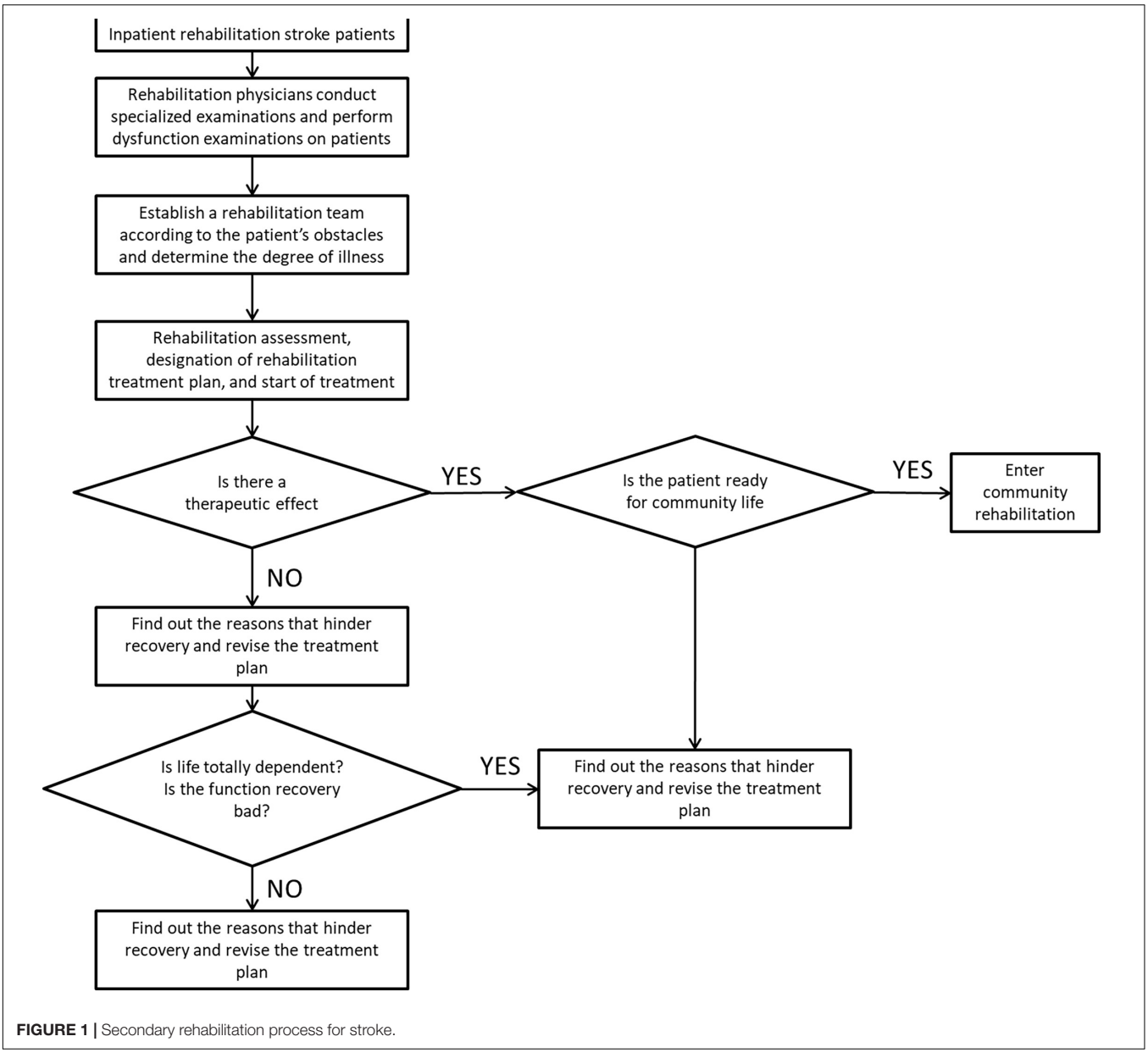


FIGURE 1 | Secondary rehabilitation process for stroke.

introduces the mean value of edge distribution features on the basis of SVM. The marginal distribution feature of LM_SVM is the first-order statistics of marginal samples, which can be expressed as:

$$\bar{\gamma} = \frac{1}{N}\sum_{i=1}^{N} y_i \left(\varphi^T(x_i)\, w + b\right) = \frac{1}{N}(Xy)^T W \tag{1}$$

LM_SVM needs to ensure to maximize the first-order statistics of the edge samples. Therefore, by introducing variables C and ξ_i, the objective function of LM_SVM is given as:

$$\min_{w,b} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\xi_i^2 - \lambda\bar{\gamma} \tag{2}$$

$$\begin{gathered} s.t. y_i\left(\varphi^T(x_i)\, w + b\right) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1 \sim N \end{gathered} \tag{3}$$

IMPROVED DEEP NEURAL NETWORK-BASED REHABILITATION MODEL

Execution Process of Rehabilitation Model Based on Improved Deep Neural Network

Aiming at the advantages of DNN and LM_SVM and the characteristics of high aliasing of stroke data, this paper proposes

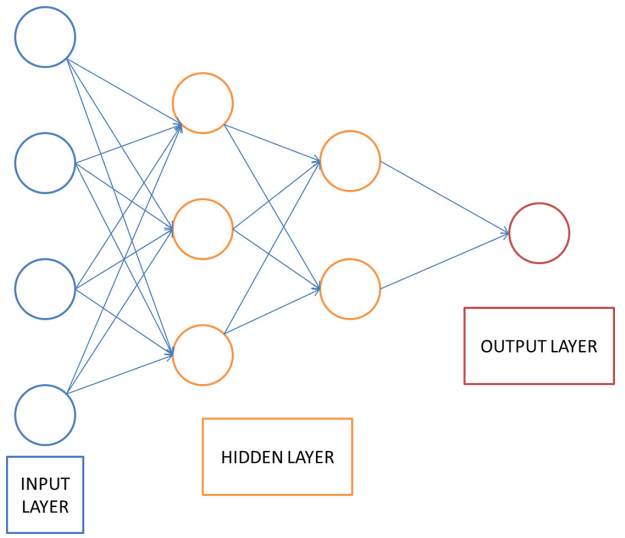


FIGURE 2 | DNN structure.

an IDNN model. This model uses DNN as the kernel function of LM_SVM to perform kernel mapping on the input data. Then the output of DNN is fed back to LM_SVM, and LM_SVM outputs the final classification result. IDNN makes up for the shortcomings of using DNN alone, so that the constructed model has better generalization. As the core part of LM_SVM, the kernel function directly affects the classification performance of the model. The execution process of the rehabilitation model based on IDNN is shown in **Figure 3**.

Improved Deep Neural Network Model

The IDNN model is mainly composed of two parts, namely DNN and LM_SVM. The DNN model of IDNN consists of an input layer and several hidden layers. Compared with other neural network models, IDNN has no output layer, and its output layer is replaced by LM_SVM. The number of hidden layers of IDNN and the number of neurons in each layer need to be determined based on experiments. LM_SVM maximizes the average value of the edge vector in the process of establishing the SVM, and separates the support vector of the SVM as much as possible. **Figure 4** depicts an IDNN structure with r hidden layers, and shows how the data changes in the model. In this model, DNN expresses the kernel mapping as an explicit function through a fully connected network. It can not only solve the kernel function selection problem of LM_SVM, but also learn more effective stroke data features, thereby improving classification performance. The output of the model is used as the input of LM_SVM, and the final classification output is completed by LM_SVM.

The general DNN model, like the perceptron, is composed of a linear relationship $z = \sum w_i x_i + b$ and an activation function $\sigma(z)$. The IDNN model proposed in this study can be regarded as a deep neural network model whose output layer is LM_SVM. Define the linear coefficient from the kth neuron of the h-1th layer to the jth neuron of the hth layer as w_{jk}^{t}. Accordingly, the bias for defining the jth neuron of the ith layer is defined as b_j^i, and

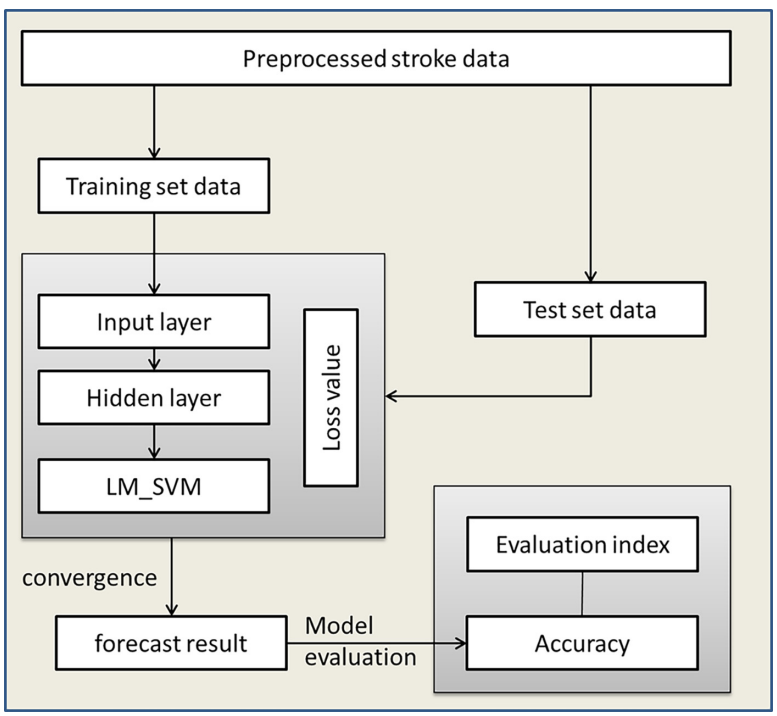


FIGURE 3 | Rehabilitation model execution process.

the output is a_j^i. For example, w_{32}^2 represents the linear coefficient from the third neuron of the second layer to the second neuron of the second layer, and the third neuron of the second layer is denoted as b_3^2. If there are a total of N neurons in the h-1th layer and n neurons in the hth layer, the linear coefficients of the lth layer can be expressed as an c^*m matrix. Correspondingly, the bias matrix of the h layer can be expressed as an c^*1 vector b^h, the output of the first layer is an c^*1 vector a^h, and the input layer's non-linearity coefficients and bias terms. Based on the above analysis, the output formula of layer h is:

$$a^h = \sigma(Z_t) = \sigma\left(W^h a^{t-1} + b^h\right) \tag{4}$$

Eq. (3) is the forward propagation expression of DNN. In the DNN model, the activation function of the output layer usually adopts softmax. Since the classification output result of DNN has an error with the actual result, it is necessary to calculate the error between the category value of the output layer and the actual value. And the error is propagated back from the output layer to the hidden layer until it propagates to the input layer. In the back propagation process, the values of various parameters are adjusted according to the error. Continue to iterate the above process until the model converges. The mean square error is often used to measure loss. The loss function of each data is as follows:

$$J = \frac{\left\|a^H - y\right\|_2^2}{2} \tag{5}$$

Where a^H represents the classification value of different categories, and y represents the actual value of different categories. Gradient descent strategy is used to solve the w and b of each layer. The loss function of the output layer can be expressed as:

$$J = \frac{\left\|a^H - y\right\|_2^2}{2} = \frac{\left\|\sigma\left(W^H a^{H-1} + b^H\right) - y\right\|_2^2}{2} \tag{6}$$

The gradient calculation formula for solving w and b is as follows:

$$\frac{\partial J}{\partial W^H} = \frac{\partial J}{\partial Z^H}\frac{\partial Z^H}{\partial W^H} = \left(a^H - y\right) \odot \sigma\left(Z^H\right)\left(Z^{H-1}\right) \tag{7}$$

$$\frac{\partial J}{\partial b^H} = \frac{\partial J}{\partial Z^H}\frac{\partial Z^H}{\partial b^H} = \left(a^H - y\right) \odot \sigma'\left(Z^H\right) \tag{8}$$

$\odot$ stands for Hadamard product. Let $\delta^H = \frac{\partial J}{\partial Z^H}$, δ^H is the gradient, the gradient of the output layer is:

$$\delta^H = \frac{\partial J}{\partial Z^H} = \frac{\partial J}{\partial Z^H}\frac{\partial Z^H}{\partial Z^{H-1}} \cdots \frac{\partial Z^{H+1}}{\partial Z^H} \tag{9}$$

From Eq. (5), the following expression is obtained:

$$Z_h = W^h a^{h-1} + b^h \tag{10}$$

Therefore, the gradients of w and b in the H-th layer can be expressed as:

$$\frac{\partial J}{\partial W^h} = \frac{\partial J}{\partial Z^h}\frac{\partial Z^h}{\partial W^h} = \delta^h\left(a^{h-1}\right)^T \tag{11}$$

$$\frac{\partial J}{\partial b^h} = \frac{\partial J}{\partial Z^h}\frac{\partial Z^h}{\partial b^h} = \delta^h \tag{12}$$

From mathematical induction, the recurrence relation of δ^h can be obtained as follows:

$$\delta^h = \delta^{h+1}\frac{\partial Z^{H+1}}{\partial Z^H} = \left(W^{h+1}\right)^T \delta^{h+1} \odot \sigma\left(Z^h\right) \tag{13}$$

In summary, the back propagation process of IDNN is summarized as follows:

Input: The total number of layers of the network, the number of neurons in each hidden layer and output layer, activation function, loss function, generation step size α, maximum number of iterations T and stopping threshold ε, and input N stroke training samples $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$.

Output: Parameters W and b in all hidden layers and output layers.

(1) Randomly initialize parameters W and b;

(2) For item from 1 to T:

(1) For h = 1 to T:

(a) a^h is set as stroke feature data x_i;

(b) For h = 2 to H, perform forward propagation calculation $a^{i,h} = \sigma\left(z^{i,h}\right) = \sigma\left(W^h \alpha^{i,h-1} + b^h\right)$;

(c) Calculate the $\delta^{i,H}$ of the output layer through the loss function;

(d) For $h = H-1$ to 2, perform back propagation calculation $\delta^{i,H} = \left(W^{h+1}\right)^T \delta^{i,h+1} \odot \sigma\left(Z^{i,h}\right)$;

(2) For h = 2 to L, update the W^h and b^h of the first layer, $W^h = W^h - \alpha\sum_{i=1}^{m}\delta^{i,h}\left(\alpha^{i,h-1}\right)^T$

(3) If all the changes of W and b are less than the iteration threshold, skip to step (3).

(3) Output the linear relationship coefficient matrix W and the bias vector b of each hidden layer and the output layer.

SIMULATION EXPERIMENT AND ANALYSIS

Experimental Data

This article uses the BCI Competition II dataset III sports imagination data set. The process of generating the data set is as follows: 280 left and right hand motor imagination EEGs of an adult female with normal physical indicators in all aspects are collected. Randomly select half of the data from the 280 times as the training set, and the remaining general data as the test set. The left and right hand imagination data in the training set each account for half. The sampling frequency of 128 Hz is used when collecting EEG, and the bandpass filter is 0.5∼30 Hz. The process of each data collection is shown in **Figure 5**.

The total time used in each experiment is 9 s. The preparation time of 2 s, at 3 s, prompts that the sample is about to perform the motor imaging task, and the screen displays the "+" symbol at the same time, and the process duration is 1 s. In the 4th second, an arrow appears on the screen, and the sample moves in the direction indicated by the arrow by dragging the feedback bar through motion imagination, and keeps it until the end of the 9th second.

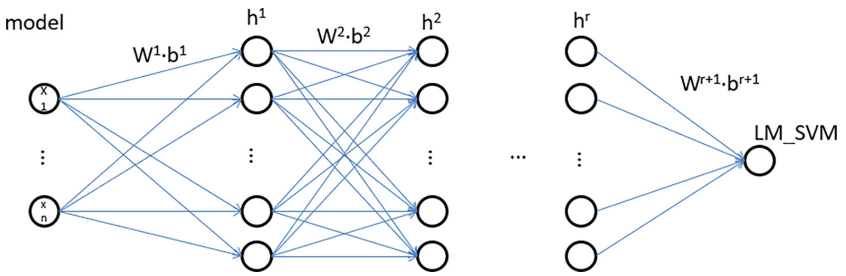


FIGURE 4 | IDNN structure.

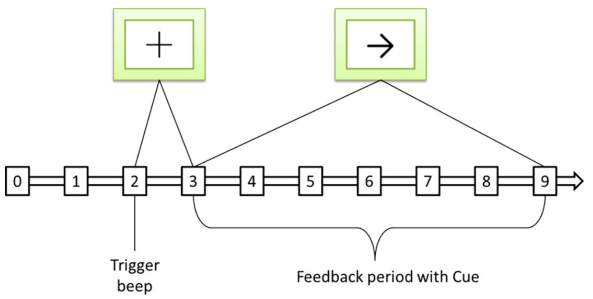


FIGURE 5 | Sequence diagram of a single motor imaging task. The symbol + represents the start signal, and the symbol → represents the directional signal.

TABLE 1 | Experimental environment.

Hardware	
Processor	RAM
Intel(R) Xeon(R) CPU E5-2650 V2@ 2.60 GHz* 4	8 GB
Software	
Operating system	Simulation environment
Microsoft Windows 7 (64-bit)	Python 3.6

**Represents the multiplication.*

Experimental Environment

This article uses the accuracy shown in Eq. (13) as the evaluation index for different classification models. Among them, TP represents the number of samples that are classified as positive and are actually positive. TN represents the number of samples that are classified as negative, but are actually negative. FP represents the number of samples classified as positive but actually negative. FN represents the number of samples classified as negative but actually positive.

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \tag{14}$$

The experimental environment is shown in **Table 1**.

Experimental Results and Analysis

Determination of Experimental Parameters

The experimental parameters that need to be confirmed during the experiment are the number of hidden layers of the network model and the number of neurons in each layer, the number of model iterations, the learning rate, and the λ value. In order to determine the structure of the network, the number of hidden layers and the number of neurons in each layer need to be determined. The experimental results of different hidden layer numbers and unit numbers are shown in **Table 2**. The number of iterations is 2,500.

The best classification accuracy is obtained when the hidden layer 1 contains 11 neural units, so the number of selected units in the first layer is 11. The classification accuracy is the best when the hidden layers 2, 3, and 5 contain 6 neural units, so the number of selected units in the first layer is 6. The best classification accuracy is obtained when the hidden layer 4 contains 8 neural units, so the number of selected units in the first layer is 8. On the other hand, as the number of hidden layers increases, the accuracy rate decreases from the original 0.9583 to 0.9386, and its changing trend is shown in **Figure 6**. This trend shows that it is not that the more hidden layers, the better the model performance. When the number of hidden layers is increased, the model construction time becomes significantly longer, but the accuracy rate begins to decrease. Therefore, considering the above factors comprehensively, this paper selects a network with two hidden layers and a combination of neurons in each layer (11, 6) as the final stroke rehabilitation model.

The difference in learning rate will also affect the performance of the network model. The learning rate range selected in this paper is 0.05–0.14, and the step size is 0.01. The experimental results of different learning rates are shown in **Table 3**. The data in **Table 3** shows that when the learning rate is 0.07, the model has the best accuracy. Therefore, the learning rate in the experiment is 0.07.

Set the learning rate to 0.78, the value range of λ is 0.01–100, and the experimental results obtained are shown in **Table 4**. According to the data in **Table 4**, the final selected value of λ is 0.1.

Performance Comparison of Different Models

The above experiment determined the number of hidden layers and the corresponding number of neurons, learning rate and *A* value. Based on these parameter settings, the structure of the network model used in this article is basically determined. In order to evaluate the effectiveness of the model used, the selected comparison models are BP neural network (Liu, 2019), DNN (Sun et al., 2021), CNN (Moon et al., 2020), LSTM

TABLE 2 | Classification accuracy of different numbers of neurons in each layer.

Hidden layer number\number of neurons	5	6	7	8	9	10	11	12
Layer 1	0.9213	0.9316	0.9320	0.9239	0.9418	0.9567	0.9583	0.9512
Layer 2	0.9522	0.9661	0.9587	0.9345	0.9465	0.9407	0.9323	0.9293
Layer 3	0.9405	0.9584	0.9509	0.9423	0.9486	0.9463	0.9394	0.9320
Layer 4	0.9373	0.9340	0.9329	0.9406	0.9234	0.9321	0.9290	0.9256
Layer 5	0.9251	0.9386	0.9318	0.9257	0.9265	0.9286	0.9255	0.9228

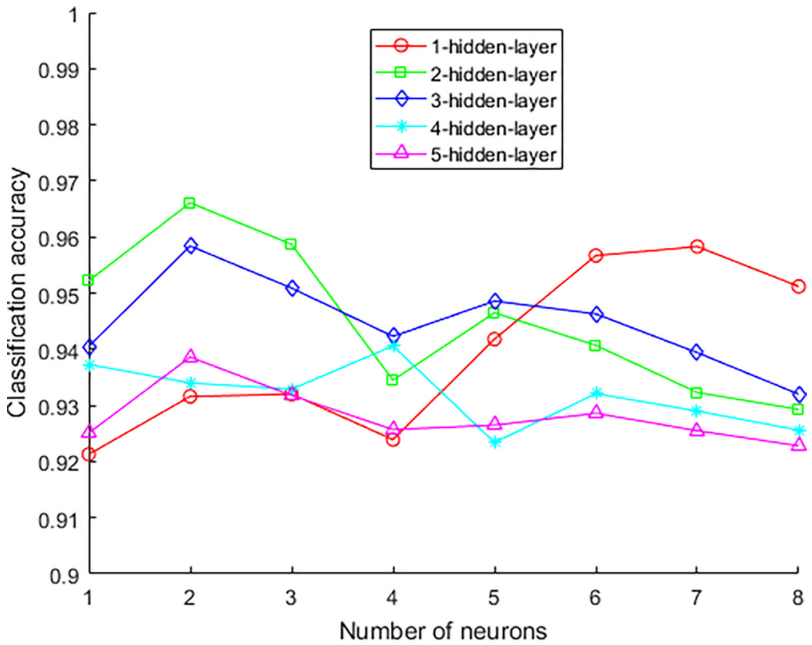


FIGURE 6 | The accuracy rate changes with the number of hidden layers.

TABLE 3 | The effect of learning rate on accuracy.

Learning rate	0.05	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14
Acc	0.9564	0.9612	0.9658	0.9451	0.9472	0.9360	0.9452	0.9401	0.9359	0.9363

TABLE 4 | The influence of λ value on accuracy.

λ	0.01	0.05	0.1	0.5	1	5	10	50	100
Acc	0.9545	0.9562	0.9660	0.9621	0.9556	0.9487	0.9506	0.9442	0.9465

TABLE 5 | Performance comparison of each model.

Index\Model	BP	CNN	DNN	LSTM	IDNN
Acc	0.9139	0.9443	0.9596	0.9611	0.9658
Time(s)	74	109	113	132	120

(Wang et al., 2018). The parameter settings of each network are the same as the corresponding references. Since rehabilitation training requires relatively high timeliness, the running time of different models is also an important indicator. **Table 5** shows the recognition accuracy and running time comparison of different models on the data set.

In terms of accuracy, the IDNN model used in this article has the highest accuracy. It shows that DNN expresses the kernel mapping as an explicit function through a fully connected network, which can not only solve the kernel function selection problem of LM_SVM, but also learn more effective stroke data features, thereby improving classification performance.

In terms of running time, the BP model takes significantly less time than other deep learning models. This fully shows that the deep learning model needs further improvement in terms of time complexity. The running time of the model used is not the least among other models of the same type, only less than LSTM. This shows that on the basis of improving the accuracy of the model used, a certain amount of time complexity has been added, but the added time is not much.

CONCLUSION

In order to better assist the rehabilitation training of stroke samples, this article uses a rehabilitation model based on deep learning technology. First, the LM_SVM classifier is introduced into the traditional DNN network. LM_SVM improves the generalization performance of the network model by optimizing the edge distribution, thereby improving the accuracy of the network's classification of EEG. Secondly, the EEG signals of stroke samples are identified based on the IDNN model. Finally, based on the recognition results, the movement of the rehabilitation equipment is controlled to assist the rehabilitation training of stroke samples. Comparative experiments show that the IDNN model proposed has a better classification accuracy for EEG. However, the time complexity has increased, but the increase is not large. In summary, how to reduce the time complexity of the model is the content of this article to continue to study in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.bbci.de/competition/iii/.

AUTHOR CONTRIBUTIONS

MG was mainly responsible for the implementation of the experiment and analysis of the results. JM was mainly responsible for the design of the entire experiment. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Scientific Research Program of Education Department of Hubei Province, China (D20184101), Higher Education Reform Project of Hubei Province, China (201707), "East Lake Scholar" of Wuhan Sports University Fund, China, and Hubei Provincial University Specialty subject group construction Special fund, China.

REFERENCES

Anderson, C. W., Stolz, E. A., and Shamsunder, S. (1995). "Discriminating mental tasks using EEG represented by AR models," in *Proceedings of the 17th International Conference of the Engineering in Medicine and Biology Society* (Montreal, QC: IEEE), 875–876.

Cortes, C., and Vapnik, V. (1995). Support vector machine. *Mach. Learn.* 20, 273–297.

Dahl, G. E., Yu, D., Deng, L., and Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. Audio, speech, and language processing. *IEEE Trans. Audio Speech Lang.* 20, 30–42.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Li, F. F. (2009). "Imagenet: a large-scale hierarchical image database[C]//computer vision and pattern recognition," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL: IEEE), 248–255.

Douiri, A., Rudd, A. G., and Wolfe, C. D. (2013). Prevalence of poststroke cognitive impairment: South London stroke register 1995-2010. *Stroke* 44, 138–145. doi: 10.1161/strokeaha.112.670844

Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402_1

Fingelkurts, A. A., Rytsala, H., and Suominen, K. (2007). Impaired functional connectivity at EEG alpha and theta frequency bands in major depression. *Hum. Brain Mapp.* 28, 247–261. doi: 10.1002/hbm.20275

Garrett, D., Peterson, D. A., and Anderson, C. W. (2003). "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," in *Proceedings of the IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 11, (Piscataway, NJ: IEEE), 141–144. doi: 10.1109/TNSRE.2003.814441

Gert, P., Gernot, R. M., Jörg, P., Hans, J. G., and Rüdiger, R. (2003). Thought'-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neurosci. Lett.* 351, 33–36. doi: 10.1016/s0304-3940(03)00947-9

Ghonchi, H., Fateh, M., Abolghasemi, V., Ferdowsi, S., and Rezvani, M. (2020). Deep recurrent-convolutional neural network for classification of simultaneous EEG-fNIRS signals. *IET Image Process.* 14, 142–153. doi: 10.1049/iet-spr.2019.0297

Gomez-Rodriguez, M., Peters, J., Hill, J., Schölkopf, B., Gharabaghi, A., and Grosse-Wentrup, M. (2011). Closing the sensorimotor loop:haptic feedback facilitates decoding of motor imagery. *J. Neural. Eng.* 8:27.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780.

Jiang, Y., Zhang, Y., Lin, C., Wu, D., and Lin, C. T. (2021). EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system. *IEEE Trans. Intell. Transp. Syst.* 22, 1752–1764.

Jiang, Y., Zhao, K., Xia, K., Xue, J., Zhou, L., Ding, Y., et al. (2019). A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation. *J. Med. Syst.* 43, 118:1–118:9.

Krebs, H. I., Hogan, N., and Aisen, M. L. (1998). Robot-aided neurorehabilitation. *IEEE Trans. Rehabil. Eng.* 6, 75–87. doi: 10.1109/86.662623

Leigh, R. H., Mijail, D. S., Gerhard, M. F., Jon, A. M., Maryam, S., Abraham, H. C., et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442, 164–171. doi: 10.1038/nature04970

Li, J., Ji, C., Yan, G., You, L., and Chen, J. (2021). An ensemble net of convolutional auto-encoder and graph auto-encoder for auto-diagnosis. *IEEE Trans. Cogn. Dev. Syst.* 13, 189–199.

Lin, Z. C., Tao, J., and Gao, Y. L. (2014). Analysis of central mechanism of cognitive training on cognitive impairment after stroke: resting-state functional magnetic resonance imaging study. *J. Int. Med. Res.* 42, 659–668.

Liu, C. Y., and Chen, C. Q. (2014). Intra- and extracranial atherosclerotic stenosis in China: epidemiology, diagnosis, treatment and risk factors. *Eur. Rev. Med. Pharmacol. Sci.* 18, 3368–3379.

Liu, L. (2019). Recognition and analysis of motor imagery EEG signal based on improved BP neural network. *IEEE Access* 7, 47794–47803. doi: 10.1109/access.2019.2910191

Moon, S., Chen, C., Hsieh, C., Wang, J. L., and Lee, J. S. (2020). Emotional EEG classification using connectivity features and convolutional neural networks. *Neural Netw.* 132, 96–107. doi: 10.1016/j.neunet.2020.08.009

Pendlebury, S. T., Chen, P. J., Welch, S. J. V., and Cuthbertson, F. C. (2015). Methodological factors in determining risk of dementia after transient ischemic attack and stroke: (III) applicability of cognitive tests. *Stroke* 46, 1494–1500. doi: 10.1161/strokeaha.115.009065

Petrosian, A., Prokhorov, D., and Homan, R. (2000). Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. *Neurocomputing* 30, 201–218. doi: 10.1016/s0925-2312(99)00126-5

Sheorajpanday, R. V. A., Nagels, G., Weeren, A. J., and De Deyn, P. (2011). Quantitative EEG in ischemic stroke: correlation with infarct volume and functional status in posterior circulation and lacunar syndromes. *Neurophysiol. Clin.* 122, 884–890. doi: 10.1016/j.clinph.2010.08.020

Sun, J., Cao, R., Zhou, M., Hussain, W., Wang, B., Xue, J., et al. (2021). A hybrid deep neural network for classification of schizophrenia using EEG Data. *Sci. Rep.* 11:4706.

Sun, J. H., Tan, L., and Yu, J. T. (2014). Post-stroke cognitive impairment: epidemiology, mechanisms and management. *Ann. Transl. Med.* 2:80.

Tokuhiro, K. (2000). Network structure and spare path capacity reduction for path failure restoration method using diversity coding. *Electron. Commun. Jpn. Part I Commun.* 83, 1–14. doi: 10.1002/1520-6424(200012)83:12<1::aid-ecja1>3.0.co;2-x

Topping, M. (2002). An overview of the development of Handy 1, a rehabilitation robot to assist the severely disabled. *J. Intell. Robot. Syst.* 34, 253–263.

Walter, J., Oyere, O., Mayowa, O., and Sonal, S. (2016). Stroke: a global response is needed. *Bull World Health Organ.* 94, 634–634A.

Wang, P., Jiang, A. M., Liu, X. F., Shang, J., and Zhang, L. (2018). LSTM-Based EEG Classification in Motor Imagery Tasks. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 26, 2086–2095. doi: 10.1109/tnsre.2018.2876129

Williams, D., Hinton, G. E., and Rumelhart, D. E. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536. doi: 10.1038/323533a0

Zandi, A. S., Javidan, M., Dumont, G. A., and Tafreshi, R. (2010). Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform. *IEEE Trans. Biomed. Eng.* 57, 1639–1651. doi: 10.1109/tbme.2010.2046417

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

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

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

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 28 October 2021
doi: 10.3389/fnins.2021.780373

Suppressing of Power Line Artifact From Electroencephalogram Measurements Using Sparsity in Frequency Domain

Jin-Lin Tan[1,2], Zhi-Feng Liang[2], Rui Zhang[2], You-Qiang Dong[1], Guang-Hui Li[1], Min Zhang[1], Hai Wang[1] and Na Xu[3]**

[1] School of Aerospace Science and Technology, Xidian University, Xi'an, China, [2] Shaanxi Aerospace Technology Application Research Institute Co., Ltd, Xi'an, China, [3] The Second Affiliated Hospital of Xiamen Medical College, Xiamen, China

Electroencephalogram (EEG) plays an important role in brain disease diagnosis and research of brain-computer interface (BCI). However, the measurements of EEG are often exposed to strong interference of power line artifact (PLA). Digital notch filters (DNFs) can be applied to remove the PLA effectively, but it also results in severe signal distortions in the time domain. To address this problem, spectrum correction (SC) based methods can be utilized. These methods estimate harmonic parameters of the PLA such that compensation signals are produced to remove the noise. In order to ensure high accuracy during harmonic parameter estimations, a novel approach is proposed in this paper. This novel approach is based on the combination of sparse representation (SR) and SC. It can deeply mine the information of PLA in the frequency domain. Firstly, a ratio-based spectrum correction (RBSC) using rectangular window is employed to make rough estimation of the harmonic parameters of PLA. Secondly, the two spectral line closest to the estimated frequency are calculated. Thirdly, the two spectral lines with high amplitudes can be utilized as input of RBSC to make finer estimations of the harmonic parameters. Finally, a compensation signal, based on the extracted harmonic parameters, is generated to suppress PLA. Numerical simulations and actual EEG signals with PLA were used to evaluate the effectiveness of the improved approach. It is verified that this approach can effectively suppress the PLA without distorting the time-domain waveform of the EEG signal.

Keywords: EEG, brain imaging, sparse representation, noise reduction, salient feature enhancement

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Zhe Yang,
Dongguan University of Technology, China
Xin Zhang,
Southwest Jiaotong University, China
Yu Wang,
Xi'an Jiaotong University, China

****Correspondence:***
Min Zhang
minzhang@xidian.edu.cn
Na Xu
xuna1313@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *21 September 2021*
Accepted: *08 October 2021*
Published: *28 October 2021*

Citation:
Tan J-L, Liang Z-F, Zhang R, Dong Y-Q, Li G-H, Zhang M, Wang H and Xu N (2021) Suppressing of Power Line Artifact From Electroencephalogram Measurements Using Sparsity in Frequency Domain. Front. Neurosci. 15:780373. doi: 10.3389/fnins.2021.780373

INTRODUCTION

In the measurement of Electroencephalogram (EEG), differences in voltages from distinct sites of the brain are recorded over a period of time (Hernández-Ronquillo et al., 2020; Sun et al., 2020). The EEG signal contains abundant information related to physiological, psychological and pathological activities of the brain. Therefore, the analysis of EEG signals is of vital importance in the clinical diagnosis and treatment of many diseases, such as Alzheimer's disease (Faiman et al., 2021), depression (Miao et al., 2021), idiopathic epilepsy and psychogenic non-epileptic seizures (Sharma G. et al., 2021). On the other hand, EEG also plays an important role in the research of brain-computer interface (BCI). BCI is referred to as a non-invasive way of human-machine

interface between the brain and exterior devices (Dagdevir and Tokmakci, 2021). BCI does not rely on conventional neurotransmission pathways (Deshpande et al., 2017). Nowadays, many brain imaging methodologies have been adopted by the researchers, such as EEG, magnetic resonance imaging (MRI) and functional MRI (fMRI; Hong et al., 2018). Among the above mediums, EEG is regarded as the most convenient one due to its mobility and lower cost.

In medical applications, owing to non-stationary characteristics of EEG features and differences among individual measurements, the analysis of EEG signals is comparatively complicated and difficult (An et al., 2021). The traditional signal analysis method based on fast Fourier transform (FFT) cannot meet the requirements of EEG signal analysis. In order to reduce excessive professional requirements for medical staff to identify EEG, deep learning (DL) based methods, such as convolutional neural network (CNN; Cao et al., 2019; Prathaban and Balasubramanian, 2021), are introduced to ensure intelligent understanding of EEG signals, especially for large scale datasets. Aiming at achieving robust recognition ability, researchers have designed novel DL neural networks with distinguished types of network architecture, input formulation, and activation function (Saegh et al., 2021). Li proposed an FFT-based deep feature learning method for EEG classification (Li and Chen, 2021). Shankar proposed a DL based epileptic seizure detection algorithm by using 2D recurrence plot images generated from EEG signals for specific brain rhythms (Shankar et al., 2021). Jana studied an efficient seizure prediction technique based on the combination of CNN and a novel technique of channel reduction (Jana and Mukherjee, 2021). According to the above studies, researchers also attached significant attentions in deriving special features from raw EEG signals because it can help improve the accuracy and efficiency of the DL neural networks.

In spite of the versatility of EEG signal in clinical applications and brain imaging science, it is reported that the salient features within the EEG signals are weak in energy. On the other hand, a variety of artifacts are likely to be incorporated during acquisitions of EEG signals. To investigate the essential physiological information corrupted by these noisy artifacts, pre-processing based on signal processing techniques is indispensable. Among the many sources of interference, the power line artifact (PLA), at the frequency of 50 or 60 Hz, is almost inevitable in "field" or mobile EEG measurements outside the lab (Leske and Dalal, 2019). The presence of PLA can significantly affect the analysis and extraction of essential features from EEG recordings. Although shielding of the environment has been recognized as an effective way to reduce PLA in the lab, it is often impractical in scenarios of natural environments. Therefore, alternative solutions, hardware based or software based, must be adopted to address this PLA problem. In comparison, the advantages of software-based solutions are more prominent because they are more flexible and easier to be realized. In state-of-the-art researches, signal processing tools, such as digital notch filters (DNFs; Piskorowski, 2013), short time Fourier transform (STFT; Huang et al., 2019), wavelet transform (WT; He et al., 2015; Huang et al., 2021; Sharma S. et al., 2021), and empirical mode decomposition (EMD; Taran et al., 2018), have been utilized to remove the PLA. However, according to theoretical investigations, the above techniques can be interpreted as digital filters with specific passing band in the frequency domain (Wu and Huang, 2004). Owing to the problems of energy leakage and picket-fence effect (PFE) in canonical Fourier transform, such kinds of digital filters will cause non-ignorable distortions in filtered results, whereas the same problems also occur in hardware-based solutions. To overcome this side effect, a feasible way is to construct a compensation signal with high precision. According to engineering experiences, PLA can be modeled as a sinusoidal component consisting of a simple harmonic wave. While, a sinusoidal component can be uniquely determined by harmonic parameters of amplitude, frequency and phase. Therefore, to retrieve the harmonic parameters become the essential task in PLA removal.

Sparse representation (SR) is a comparatively new development of signal expansion (Li and Chen, 2008). It expresses an input signal in terms of linear combinations of atoms from a dictionary. It has achieved tremendous successes in various engineering applications (Wang et al., 2016; Yang et al., 2020; Zhang et al., 2021). Conventional SRs are implemented using iterative algorithms and may require large computational resources (Cao et al., 2021; Chen et al., 2021; An et al., 2022). If the efficiency of SR algorithm can be improved, it will effectively expand the scope of its use in medical engineering (Yang et al., 2019; Collazos-Huertas et al., 2020; Cury et al., 2020). In this paper, to overcome the signal distortion effect in PLA removal of EEG signals, we propose a novel approach based on the combination of SR and spectrum correction (SC). SC methods can identify the harmonic parameters based on the information in the frequency domain. It will be shown in this paper that for a sinusoidal component, the SC can be regarded as a numerical algorithm to realize SR *via* a non-iterative way. In ratio-based spectrum correction (RBSC) algorithms, the performance of harmonic parameter extraction is determined by signal-to-noise ratios of baseline spectral lines (BSSLs). Taking this phenomenon into account, a dual-step correcting algorithm (DSCA) is put forward. Firstly, a RBSC, using the rectangular window, is employed for preliminary identification of harmonic parameters. Secondly, two BSSLs with high SNR are calculated. Thirdly, another RBSC is employed to estimate the harmonic parameters with higher precision. On the basis of the estimated harmonic parameters, a compensation signal is reconstructed. PLA removal by subtracting the compensation signal from the original EEG signal can avoid the side effect of signal distortion. Analysis using numerical simulation and measured signal shows that the proposed method is effective, efficient, and can suppress waveform distortion.

MATERIALS AND METHODS

Mechanism of Signal Distortion in Electroencephalogram Pre-processing

To ensure high performance of BCI, noises in EEG signals should be suppressed using pre-processing based on hardware-based

or software-based solutions. However, severe signal distortions may be produced by such pre-processing. The theory of Fourier transform can be used to explain the reason of this side effect. An EEG signal$x(t)$, corrupted by the PLA, can be expressed as$x(t) = eeg(t) + pla(t)$, where $eeg(t)$ and $pla(t)$ stands for the salient EEG features and PLA, respectively. In most cases, the PLA component can be modeled as a sinusoidal component, expressed as$pla(t) = A_{pla} \cdot \cos(2\pi f_{pla} t + \varphi_{pla})$. These harmonic parameters (A_{pla},f_{pla},φ_{pla}) are crucial to determine the waveform of PLA. In actual analysis of a dynamic process, the digitization of a physical quantity requires the signal to have a finite sampling frequency (f_s) and a finite sampling number (N) (**Figure 1A**). That is to say, after digitization, $x(t)$ should be expressed as $x[n] = x((n-1)\cdot\Delta t)$, where $n = 0, 1, \ldots, N-1$ and the spatial interval is denoted as $\Delta t = 1/f_s$. The FFT converts an input signal from the time domain into a counterpart in the frequency domain. In numerical implementations of FFT, a cluster of sinusoidal waves are used to decompose the input signals.

$$X[\ell] = FFT\{x[n]\}, \tag{1}$$

where $X[\ell]$ ($\ell = 0, 1 \ldots, N-1$) represents the complex-valued coefficient at the frequency of $(\ell-1)\cdot f_s/N$. The FFT spectrum is also a sampling of actual Fourier spectrum of $x(t)$. If the frequency f_{pla} does not belong to the set $\{(\ell-1)\cdot f_s/N |\ \ell = 0, 1 \ldots, N-1\}$, the problem of energy leakage and PFE occur (Chen et al., 2019). In such circumstance, the FFT spectrum of $\{pla[n]\}$ is dense in frequency domain (**Figure 1B**). The energy of the noisy component $\{pla[n]\}$ leaks across the entire frequency domain. A simple harmonic wave in time domain becomes a broad band signal in the frequency domain. Therefore, the harmonic parameters cannot be estimated using any single spectral line in the FFT spectrum (**Figure 1C**).

Wavelet transform and EMD are important developments of the classical FFT. However, they are essentially digital filters with specific passing band in the frequency domain (**Figure 1D**). By using them, only a part of the PLA spectrum can be suppressed. The residual contents after digital filtering still leads to side effect of signal distortion. This phenomenon is similar to Gibbs phenomenon in Fourier analysis. Therefore, it can also be summarized as a pseudo-Gibbs phenomenon (PGF) in multi-scale analysis.

Fundamentals of Sparse Representation

Sparse representation has become a hot research topic in the field of signal processing during the past two decades. SR aims at expressing an input signal x as a linear combination of atoms from an over-complete dictionary Φ.

$$x = \Phi c = \begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_M \end{bmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{pmatrix}, \tag{2}$$

where $\varphi_i \in \Phi$ ($i = 1, 2, \ldots, M$) and M is the number of atoms in Φ. For EEG signal analysis, the variables x and φ_i are supposed to be column vectors of dimension K. The dimension of Φ in matrix form is $K \times M$. Different from a basis, the condition $M \times K$ should be satisfied for an SR dictionary Φ. To satisfy the demand of sparsity, most coefficients in the set $\{c_i\}$ are approximate or equal to zero.

To obtain a sparse solution $\hat{c}$, we need to solve the following optimization problem.

$$\hat{c} = argmin \parallel c \parallel_0, \quad s.t. \ \ x = \Phi c \tag{3}$$

where the L0-norm $||c||_0$ calculates the number of non-zero terms in $\hat{c}$. However, this problem has proven to be NP-hard. To deal with this problem, the L1-norm $||c||_1 = \sum_{i=1}^{K} |c_i|$ can be introduced to derive a feasible solution. Therefore, the problem in Eq. (3) is formulated as below.

$$\hat{c} = argmin \parallel c \parallel_1, \quad s.t. \ \ x = \Phi c \tag{4}$$

In practical applications, owing to the existence of measurement noises, the constraint condition $x = \Phi c$ is usually too strict that may results in no solution. A relaxed L2-norm constraint, allowing a maximum error of ε, is more convenient.

$$\hat{c} = argmin \parallel c \parallel_1, \quad s.t. \quad \parallel x - \Phi c \parallel_2 \leq \varepsilon \tag{5}$$

To solve the above optimization problem, iterative algorithms, such as matching pursuit and basis pursuit, can be utilized. From the above argument, the solution of $\hat{c}$ depends on the choosing of a proper dictionary Φ. Therefore, a critical problem of SR in engineering applications lies in the selection a proper over-complete dictionary.

Spectrum Correction for Sinusoidal Component

Fast Fourier transform is extensively applied to investigate the component of $x(t) = A_c \cdot \cos(2\pi f_c t + \varphi_c)$, which appears as a simple harmonic wave in a dynamic process. According to the fundamentals of digital signal processing, if the condition $f_s/f_c \in N^+$ is satisfied, the information of the PLA noise can be uniquely revealed by one spectral line, whose frequency is exactlyf_c, in the FFT spectrum. However, this condition is difficult to meet in most cases because the value of f_c can be a complicated decimal. In such circumstances, SCs, based on the information in FFT spectra, should be adopted. In this section, we take the rectangular window function as an example to illustrate the SC algorithm.

In FFT analysis, the spectral resolution is defined as $\Delta f = 1/(N\cdot\Delta t)$. Spectral lines are uniformly spaced by one single spectral resolution in the frequency axis. According to the algorithm of RBSC, a pair of spectral lines whose frequencies differ by one spectral resolution are utilized to estimate the harmonic parameters. For ease of argument, we name the two spectral lines as BSSLs. The frequencies of BSSL are denoted by f_l and f_r, and they are belonging to the spectral grid set $\{(\ell-1)\cdot f_s/N |\ \ell = 0, 1 \ldots, N-1\}$. To estimate harmonic parameters of a sinusoidal component at the frequency of f_c, the following relationship is satisfied.

$$f_c - 0.5 \leq f_l < f_c < f_r \leq f_c + 0.5 \tag{6}$$

The concept of normalized index of spectral line is used to indicate the frequency of a spectral line in FFT spectrum. Let

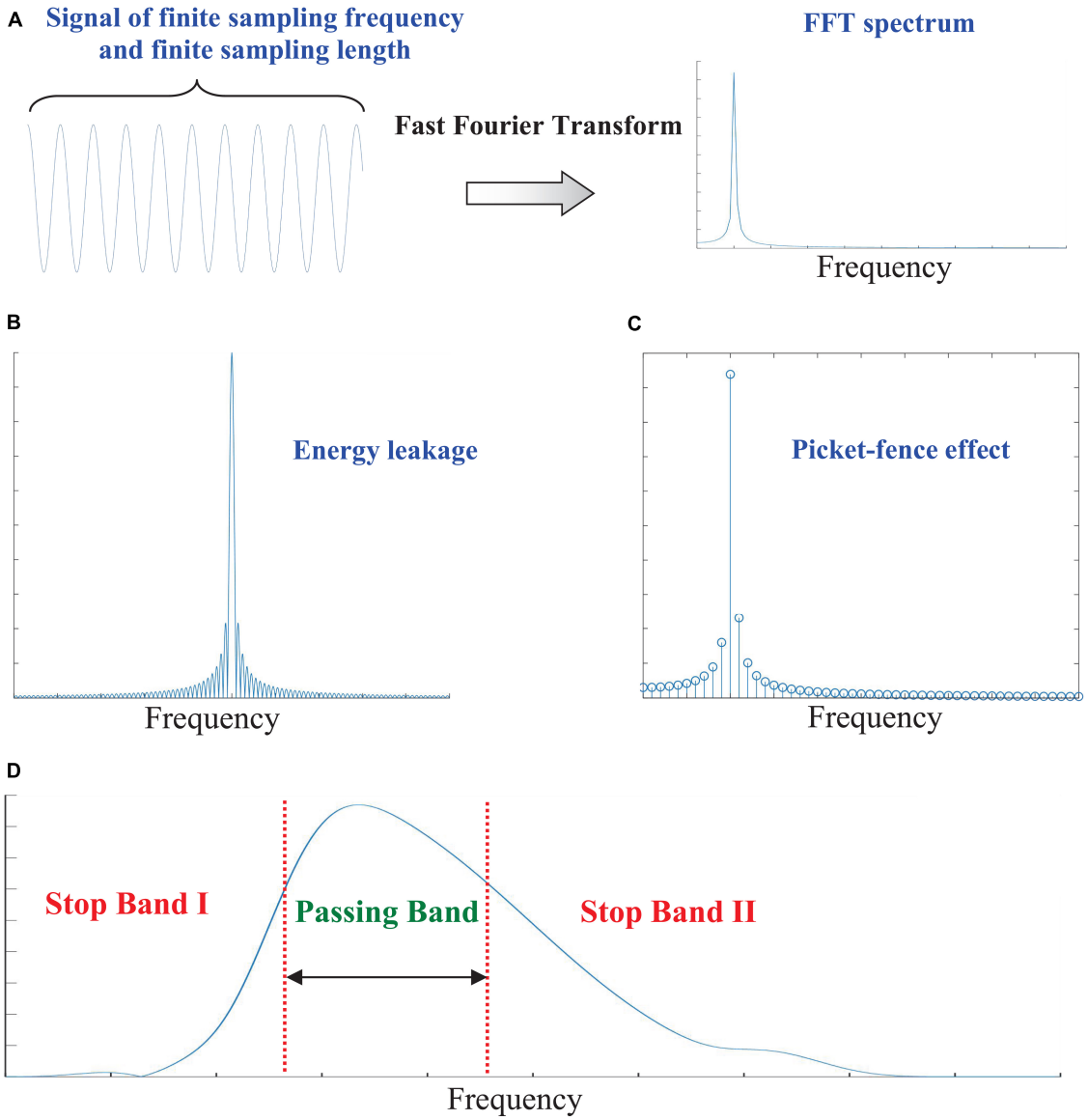


FIGURE 1 | (A) Digitization of dynamic processes with finite sampling frequency and a finite sampling number; **(B)** FFT spectrum of a sinusoidal component; **(C)** the energy leakage problem; **(D)** the picket-fence effect (PFE) and the filtering characteristics of a digital filters from Daubechies wavelet basis (Mallat, 1989; Chen et al., 2012).

Y_k be the complex number associated with the spectral line whose frequency is $(k-1)\cdot\Delta f$, y_k and $\arg(Y_k)$ can be used to represent amplitude and argument of the Y_k. **Figure 2A** shows the amplitudes (y_l and y_r) of BSSLs in RBSC. The normalized frequency shift is calculated by $\Delta k = (f_c - f_l)/\Delta f$. In RBSC, the ratio between y_l and y_r can be used to calculate Δk conveniently, which is implemented based on the information fusion of BSSLs. An indicator of characteristic ratio for the SC can be defined as $R = y_l/y_r$. Using the characteristic ratio, the normalized error Δk can be calculated as below.

$$\Delta k = -\frac{1}{1+R} = -\frac{y_k}{y_k + y_{k+1}} \tag{7}$$

Then the estimated harmonic parameters ($\tilde{A}_c$, $\tilde{f}_c$, and $\tilde{\varphi}_c$) of the investigated sinusoidal component can be derived as below.

$$\begin{cases} \tilde{A}_c = \frac{\pi\cdot\Delta k\cdot y_k}{\sin(\pi\cdot\Delta k)} \\ \tilde{f}_c = (k+\Delta k)\cdot\Delta f \\ \tilde{\varphi}_c = \arg(Y_k) - \pi\cdot\Delta k \end{cases} \tag{8}$$

The sinusoidal component is supposed to be produced using the estimated harmonic information. In summary, the RBSC produces a harmonic atom that is very similar to the original PLA. This harmonic atom can be treated as a special dictionary with $M = 1 << K$. This is the relationship between SC and SR for sinusoidal component analysis.

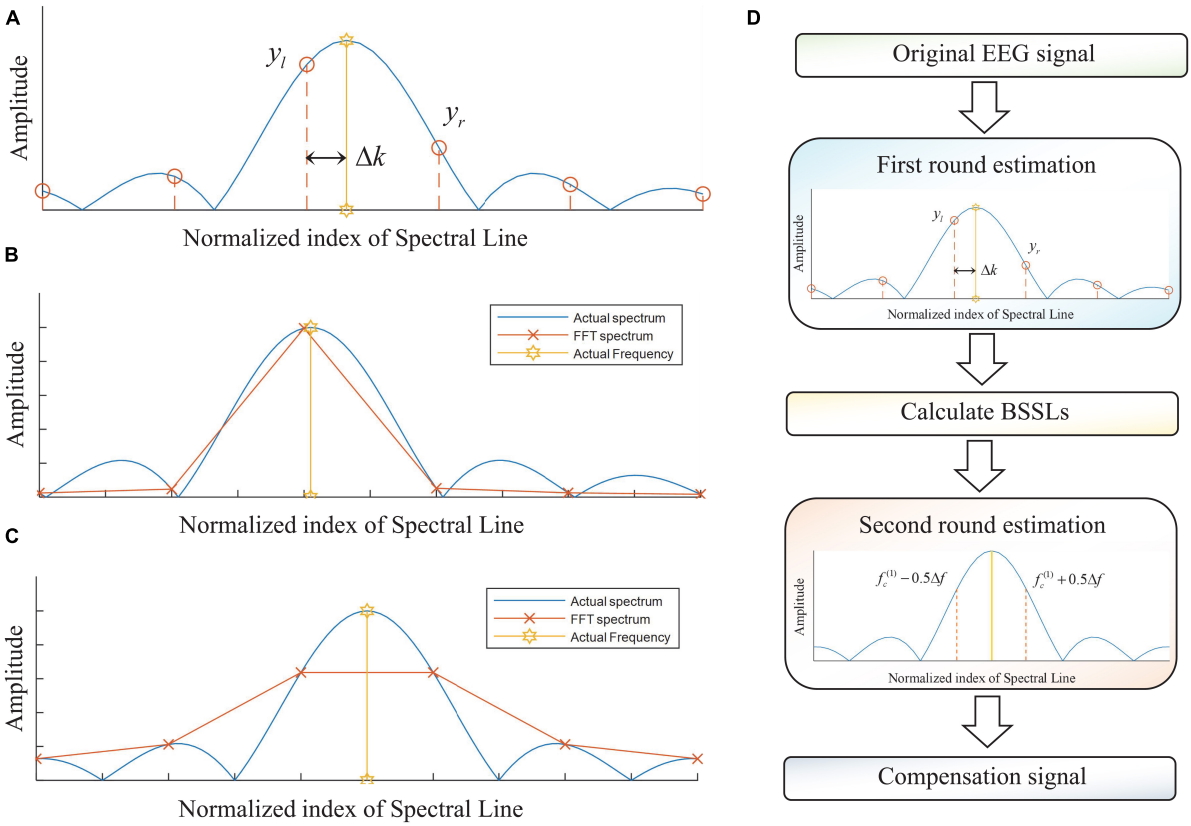


FIGURE 2 | (A) Mathematical model for the RBSC; **(B)** RBSC for small normalized frequency error; **(C)** RBS for normalized frequency error $\Delta k = 1/2$; and **(D)** the flowchart of the DBSC.

The Proposed Method

The RBSC is an effective method to estimate harmonic parameters of a sinusoidal component. It utilizes information of BSSLs in FFT spectrum to estimate the harmonic parameters. In actual dynamic measurements, the information of BSSLs is also affected by noises. Therefore, high SNR are important to ensure estimations of high precision. As the value of Δk approaches zero, the SNR of y_r reduces significantly if measurement noise exists (**Figure 2B**). As a result, significant estimation errors can be produced for the harmonic parameters. As a compromise, we expect Δk to be approximately 1/2 such that both of y_l and y_r are of high amplitudes to resist the measurement noises (**Figure 2C**). To address this problem, a DSCA is proposed. The key idea of DSCA is the numerical implementation of spectral correction using BSSLs of high amplitudes. The flowchart of the DBSC is shown in **Figure 2D**.

Step 1. Apply FFT for the input signal $x(t) = A_c \cdot \cos(2\pi f_c t + \phi_c)$.

$$x(t) \xrightarrow{FFT} X(k) \tag{9}$$

Step 2. First round estimation of harmonic parameter in $x(t)$ using RBSC.

$$X(k) \xrightarrow{RBSC} A_c^{(1)},\ f_c^{(1)},\ \varphi_c^{(1)} \tag{10}$$

Step 3. Calculate BSSLs at frequencies of $f_c^{(1)} - 0.5 \cdot \Delta f$ and $f_c^{(1)} + 0.5 \cdot \Delta f$ based on $f_c^{(1)}$ and the FFT spectrum.

$$\begin{cases} f_c^{(1)} \\ X(k) \end{cases} \longrightarrow \begin{cases} f_c^{(1)} - 0.5 \cdot \Delta f \\ f_c^{(1)} + 0.5 \cdot \Delta f \end{cases} \tag{11}$$

Step 4. Second round estimation of the harmonic parameters $(\tilde{A}_c,\ \tilde{f}_c,\ \tilde{\varphi}_c)$ using RBSC.

$$\begin{cases} f_c^{(1)} - 0.5 \cdot \Delta f \\ f_c^{(1)} + 0.5 \cdot \Delta f \end{cases} \xrightarrow{RBSC} A_c^{(2)},\ f_c^{(2)},\ \varphi_c^{(2)} \tag{12}$$

Step 5. The harmonic parameters are corrected as below.

$$\begin{cases} \tilde{A}_c = A_c^{(1)} \\ \tilde{f}_c = f_c^{(2)} \\ \tilde{\phi}_c = \varphi_c^{(2)} \end{cases} \tag{13}$$

Step 6. A compensation signal can be produced using the corrected harmonic information.

$$\tilde{x}(t) = \tilde{A}_c^{(1)} \cdot \cos(2\pi \tilde{f}_c^{(2)} t + \tilde{\varphi}_c^{(2)}) \tag{14}$$

According to the above algorithm, no iteration is employed. The original sinusoidal component is retrieved with errors

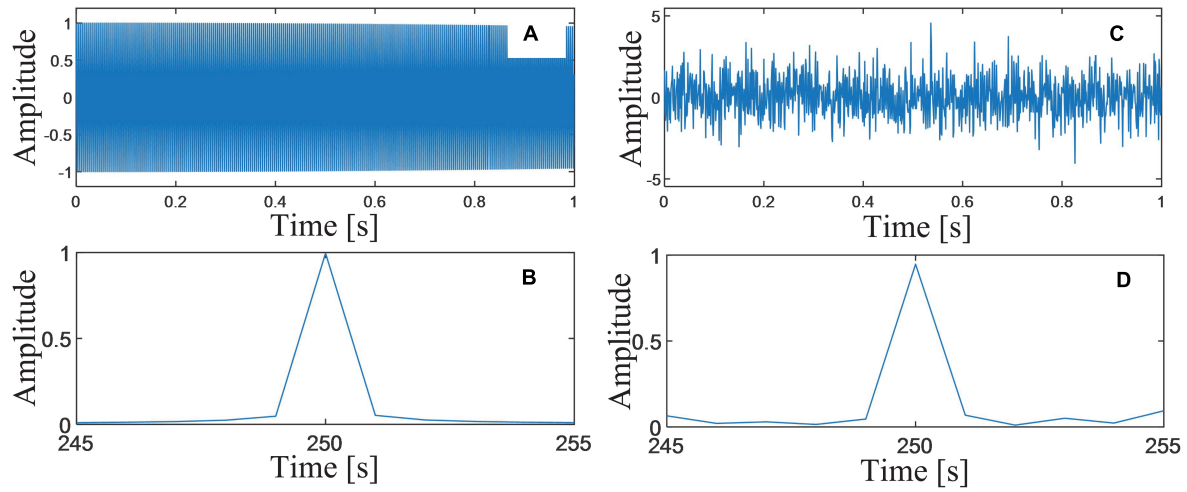


FIGURE 3 | (A) Waveform of the simulated PLA; **(B)** FFT spectrum of the simulated PLA; **(C)** the simulated PLA with noises (SNR = 0dB); and **(D)** FFT spectrum of the simulated PLA with noises.

$\| x(t) - \tilde{x}(t) \| > 0$. Only one atom, learned by the spectral corrections, is used to reconstruct $x(t)$. This can be seen as a special form of SR for sinusoidal component.

RESULTS

Numerical Simulations

To verify the enhancement of the proposed DSCA, numerical simulations are employed in this sub-section. A signal, consisting of PLA ($pla(t)$, **Figure 3A**) and noises ($n(t)$), is simulated. Without loss of generality, the harmonic parameters are set as $A_{pla} = 1$, $f_{pla} = 250.05Hz$, and $\varphi_{pla} = 0$. A white Gaussian noise $n(t)$ is added to the simulated signal, and the SNR is set as 0dB (**Figure 3C**). Both of the sampling frequency and the sampling number are set as 1,000. The spectral resolution of the signal in spectral domain is calculated as $\Delta f = 1$ (**Figures 3B,D**). In FFT spectrum of $pla(t)$, the frequencies of BSSLs are 250 and 251 Hz. Their amplitudes are $y_l = 0.996$ and $y_r = 0.052$. The normalized frequency shift is $\Delta k = 0.05$, which is very closed to zero. The harmonic parameters are corrected as $A_{pla} = 1$, $f_{pla} = 250.05Hz$, and $\phi_{pla} = 0$. The estimation accuracy is perfect when there is no noise. While in the spectrum of $s(t)$, $y_l = 0.947$ and $y_r = 0.068$. The harmonic parameters are corrected as $A_{pla} = 0.9539$, $f_{pla} = 250.067Hz$, and $\phi_{pla} = -0.010$. It can be seen significant estimation errors occur due to noises.

For the proposed DSCA, the first-round estimation of the frequency is $f_c^{(1)} = 250.067$. The frequencies of BSSLs are $f_l = 249.567$ and $f_r = 255.067$. The harmonic parameters are corrected as $A_{pla} = 1.073, f_{pla} = 250.040\,Hz$, and $\varphi_{pla} = -0.006$. It can be seen that the estimation errors of the frequency and the phase are reduced significantly, while estimation error of the phase increases. The same rule of error has also been verified in other independent repeated experiments. That is the reason why we use the estimated harmonic parameters in different rounds of estimations.

$$s(t) = pla(t) + n(t) \tag{15}$$

Case Study of Noisy Electroencephalogram Signals

In this sub-section, actual measurements of EEG signals are employed to further verify the performance of the proposed DSCA. The EEG datasets, we used in this paper, were made available to the public by Neurology & Sleep Centre, Hauz Khas, New Delhi. The employed datasets contain EEG records from clinical studies concerning epilepsy. A record of EEG time series is selected from the datasets (**Figure 4A**). The sampling frequency is 200 Hz, and the sampling number is 1,000. In acquisition of the EEG time series, a band-pass digital filter with the passing band of $[0.5, 70]Hz$ was applied. To simulate the PLA, an additive noisy component is synthesized with the parameters $A_{pla} = 50$, $f_{pla} = 250.05\,Hz$, and $\varphi_{pla} = 0$(**Figure 4B**). The FFT spectra of the two components are shown in **Figure 4C**. A noisy EEG signal is formed by directly superimposing the two components. The SNR of the investigated signal is 3.430 dB.

The proposed DSCA is applied to remove the PLA. The harmonic parameters are corrected as $A_{pla} = 50.006$, $f_{pla} = 250.050\,Hz$, and $\varphi_{pla} = -0.023$. Using these estimated parameters, the compensation signal can be reconstructed and subtracted from the synthesized signal. The denoised signal is plotted in **Figure 4D**. From the appearances of the two waveforms, they are very similar. The correlation coefficient of the two signals are computed as 1.00, indicating that the filtering result is satisfactory. The FFT spectra of the denoised signal and the original noisy EEG signal are shown in **Figure 4E**. The difference only exists in a very narrow frequency band around 50 Hz.

As comparison, a DNF based method is also used to process the noisy signal (Chen et al., 2012). The denoised signal is shown

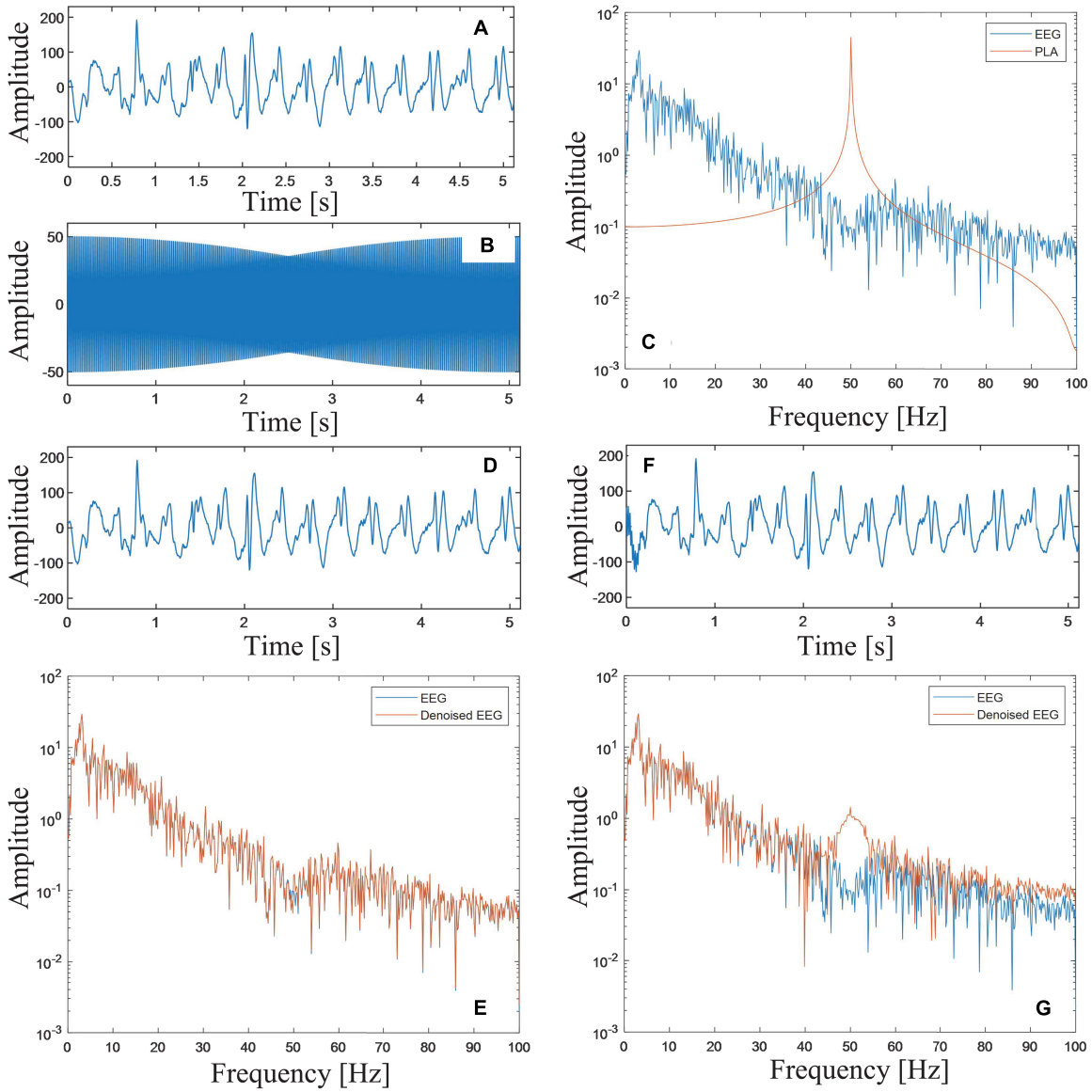


FIGURE 4 | (A) Time domain waveform of the EEG signal; **(B)** the simulated PLA noise; **(C)** FFT spectra of the EEG signal and the PLA noise; **(D)** the denoised signal based on the proposed DSCA; **(E)** FFT spectra of the EEG signal and the filtered signal using DSCA; **(F)** the denoised signal based on DNF method; and **(G)** FFT spectra of the EEG signal and the filtered signal using DNF method.

in **Figure 4F**. In the time domain waveform, there is a severe distortion on the left side of the filtered signal. The correlation coefficient of the two signals are computed as 0.997. While, from the frequency domain (**Figure 4G**), differences between the two signals increase significantly near 50 Hz. The above comparisons show that the proposed DSCA algorithm in this paper outperforms DNF in PLA removal.

DISCUSSION

The problems of energy leakage and PFE are inevitable for harmonic components whose digital samplings do not satisfy the full period sampling condition (Mewett et al., 2001). There is a view that these two side effects can be prevented by actively adjusting the sampling parameters. Because a variation of $\pm$ 2 Hz is likely to occur in actual power systems (Singhal et al., 2020), the parameter f_{pla} is an unknown variable. It is not practical to design DNFs perfectly suitable for all possible PLA noises. On the other hand, increasing the sampling length of EEG signal can improve the accuracy of harmonic parameter estimation. However, the process of brain electrical activity corresponding to the collected signals cannot be repeated, so we can only analyze and process the signals of limited length.

Wavelet transform and EMD outperform FFT by providing time-frequency representation of EEG signal. However, in either

case, the above side effect is still unavoidable. Because no matter how they decompose the signal, they can only suppress part of the PLA, and the rest will still cause signal distortion. In principle, RBSC uses local spectral information to estimate the overall harmonic composition, so it can avoid the side effect.

The method proposed in this paper is an improvement of the classical RBSC. It not only improves the robustness of the algorithm in parameter correction, but also considers the efficiency of the algorithm. The complexity of the algorithm has only a small increase compared with the classical method. The method proposed in this paper is especially suitable for the case that the value of Δk is very small. When the value of this parameter is close to half the frequency resolution, the performance of the two methods is close. In general, the method proposed in this paper can be better applied to the complex measurement environment. However, in any case, the method in this paper has no less accuracy than the original RBSC in statistics. So we recommend that DSCA can completely replace the original RBSC.

Ratio-based spectrum correction only utilize two BSSLs to estimate harmonic parameters, and it can balance the efficiency and accuracy of the algorithm. In the literature, there have been some methods for harmonic parameter estimation using multiple BSSLs. These studies also show that their harmonic parameter estimation accuracy is indeed better than that of RBSC in the absence of noise. But this is not achievable in clinical applications. Therefore, the DSCA is proposed based on the RBSC. The calculation time of RBSC mainly includes the FFT transform of the signal and several complex arithmetic operations. The DSCA proposed in this paper mainly includes two RBSCs, so the efficiency is also very high.

For analysis of a simple harmonic component, the spectral correction is very similar with the spare representation theory in mathematical principle. A remarkable advantage is that it does not require a predetermined dictionary. The algorithm can generate a skinny dictionary containing only one atom to represent the actual sinusoidal component. Because the dictionary is not redundant, no iterative algorithms is needed. We can call the spectral correction as implicit sparse representation (ISR) for simple harmonic component analysis. The current research shows that the results obtained by using SR based on iterative numerical algorithm are very similar to those based on spectral correction method. Therefore, in the case of harmonic analysis, a method based on SC may be preferentially employed.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the ictal repository, accession number ictal1 (https://www.researchgate.net/publication/308719109_EEG_Epilepsy_Datasets).

AUTHOR CONTRIBUTIONS

MZ conceived and designed the classification method. Z-FL, RZ, and NX performed the experiment. Z-FL, RZ, Y-QD, and G-HL analyzed the data and wrote the manuscript. MZ and NX reviewed and edited the manuscript. All authors read and approved the manuscript.

FUNDING

This research is financially supported by the Fundamental Research Funds for the Central Universities (XJS191305).

REFERENCES

An, F. P., Ma, X. M., and Bai, L. (2022). Image fusion algorithm based on unsupervised deep learning-optimized sparse representation. *Biomed. Signal Process. Control* 71, (Pt B):103140**. doi: 10.1016/j.bspc.2021.103140

An, P. F., Yuan, Z. Y., and Zhao, J. H. (2021). Unsupervised multi-subepoch feature learning and hierarchical classification for EEG-based sleep staging. *Expert Syst. Appl.* 186:18115759. doi: 10.1016/j.eswa.2021.115759

Cao, J. X., Liu, S. J., Liu, H. Q., and Zhang, K. (2021). MRI reconstruction based on Bayesian group sparse representation. *Signal Process.* 187:108151. doi: 10.1016/j.sigpro.2021.108151

Cao, X. C., Chen, B. Q., Yao, B., and He, W. P. (2019). Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification. *Comput. Ind.* 106, 71–84. doi: 10.1016/j.compind.2018.12.018

Chen, B. Q., Lan, Q. X., Li, Y., Zhuang, S. Q., and Cao, X. C. (2019). Enhancement of fault feature extraction from displacement signals by suppressing severe end distortions via sinusoidal wave reduction. *Energies* 12:3536. doi: 10.3390/en12183536

Chen, B. Q., Li, Y., Cao, X. C., Suen, V., and Zhuang, S. (2021). Recovery of undersampled force measurement from high-speed milling process using approximate sparsity in frequency domain. *Measurement* 175:109143. doi: 10.1016/j.measurement.2021.109143

Chen, B. Q., Zhang, Z. S., Sun, C., Li, B., Zi, Y., and He, Z. (2012). Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors. *Mech. Syst. Signal Process.* 33, 275–298. doi: 10.1016/j.ymssp.2012.07.007

Collazos-Huertas, D., Caicedo-Acosta, J., Castaño-Duque, G. A., and Acosta-Medina, C. D. (2020). Enhanced multiple instance representation using time-frequency atoms in motor imagery classification. *Front. Neurosci.* 14:155. doi: 10.3389/fnins.2020.00155

Cury, C., Maurel, P., Gribonval, R., and Barillot, C. (2020). A sparse EEG-Informed fMRI model for Hybrid EEG-fMRI neurofeedback prediction. *Front. Neurosci.* 13:1451. doi: 10.3389/fnins.2019.01451

Dagdevir, E., and Tokmakci, M. (2021). Optimization of preprocessing stage in EEG based BCI systems in terms of accuracy and timing cost. *Biomed. Signal Process. Control* 67:102548. doi: 10.1016/j.bspc.2021.102548

Deshpande, G., Rangaprakash, D., Oeding, L., Cichocki, A., and Hu, X. P. (2017). A new generation of brain-computer interfaces driven by discovery of latent EEG-fMRI linkages using tensor decomposition. *Front. Neurosci.* 11:246. doi: 10.3389/fnins.2017.00246

Faiman, I., Smith, S., Hodsoll, J., Young, A. H., and Shotbolt, P. (2021). Resting-state EEG for the diagnosis of idiopathic epilepsy and psychogenic nonepileptic seizures: a systematic review. *Epilepsy Behavior.* 121:108047. doi: 10.1016/j.yebeh.2021.108047

He, W. P., Zi, Y. Y., Chen, B. Q., Wu, F., and He, Z. J. (2015). Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform. *Mech. Syst. Signal Process.* 54–55, 457–480. doi: 10.1016/j.ymssp.2014.09.007

Hernández-Ronquillo, L., Thorpe, L., Dash, D., Hussein, T., Hunter, G., Waterhouse, K., et al. (2020). Diagnostic accuracy of the ambulatory EEG vs.

routine EEG for first single unprovoked seizures and seizure recurrence: the DX-seizure study. *Front. Neurol* 11:223. doi: 10.3389/fneur.2020.00223

Hong, K.-S., Khan, M. J., and Hong, M. J. (2018). Feature extraction and classification methods for Hybrid fNIRS-EEG brain-computer interfaces. *Front. Hum. Neurosci.* 12:246. doi: 10.3389/fnhum.2018.00246

Huang, J. S., Chen, B. Q., Li, Y., and Sun, W. F. (2021). Fractal geometry of wavelet decomposition in mechanical signature analysis. *Measurement* 173:108571. doi: 10.1016/j.measurement.2020.108571

Huang, J. S., Chen, B. Q., Yao, B., and He, W. (2019). ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access* 7, 92871–92880. doi: 10.1109/ACCESS.2019.2928017

Jana, R., and Mukherjee, I. (2021). Deep learning based efficient epileptic seizure prediction with EEG channel optimization. *Biomed. Signal Process. Control* 68:102767. doi: 10.1016/j.bspc.2021.102767

Leske, S., and Dalal, S. S. (2019). Reducing power line noise in EEG and MEG data via spectrum interpolation. *NeuroImage* 189, 763–776. doi: 10.1016/j.neuroimage.2019.01.026

Li, M. Y., and Chen, W. Z. (2021). FFT-based deep feature learning method for EEG classification. *Biomed. Signal Process. Control* 66:102492. doi: 10.1016/j.bspc.2021.102492

Li, Y. F., and Chen, K. F. (2008). Eliminating the picket fence effect of the fast Fourier transform. *Comput. Phys. Commun.* 178, 486–491. doi: 10.1016/j.cpc.2007.11.005

Mallat, S. G. (1989). A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11, 674–693. doi: 10.1109/34.192463

Mewett, D. T., Nazeran, H., and Reynolds, K. J. (2001). "Removing power line noise from recorded EMG," in *Proceedings of the 23rd Annual International Conference of the IEEE*, Istanbul, doi: 10.1109/IEMBS.2001.1017205

Miao, Y., Jurica, P., Struzik, Z. R., Hitomi, T., Kinoshita, A., Takahara, Y., et al. (2021). Dynamic theta/beta ratio of clinical EEG in Alzheimer's disease. *J. Neurosc. Methods* 359:109219. doi: 10.1016/j.jneumeth.2021.109219

Piskorowski, J. (2013). Time-efficient removal of power-line noise from EMG signals using IIR notch filters with non-zero initial conditions. *Biocybern. Biomed. Eng.* 33, 171–178. doi: 10.1016/j.bbe.2013.07.006

Prathaban, B. P., and Balasubramanian, R. (2021). Dynamic learning framework for epileptic seizure prediction using sparsity based EEG reconstruction with optimized CNN classifier. *Expert Syst. Appl.* 170:114533. doi: 10.1016/j.eswa.2020.114533

Saegh, A. A., Dawwd, S. A., and Jabbar, J. M. A. (2021). Deep learning for motor imagery EEG-based classification: a review. *Biomed. Signal Process. Control* 63:102172. doi: 10.1016/j.bspc.2020.102172

Shankar, A., Khaing, H. K., Danpapat, S., and Barma, S. (2021). Analysis of epileptic seizures based on EEG using recurrence plot images and deep learning. *Biomed. Signal Process. Control* 69:102854. doi: 10.1016/j.bspc.2021.102854

Sharma, G., Paashar, A., and Joshi, A. M. (2021). DepHNN: a novel hybrid neural network for electroencephalogram (EEG)-based screening of depression. *Biomed. Signal Process. Control* 66:102393. doi: 10.1016/j.bspc.2020.102393

Sharma, S., Khare, S. K., Bajaj, V., and Ansari, I. A. (2021). Improving the separability of drowsiness and alert EEG signals using analytic form of wavelet transform. *Appl. Acoust.* 181:108164. doi: 10.1016/j.apacoust.2021.108164

Singhal, A., Singh, P., Fatimah, B., and Pachori, R. B. (2020). An efficient removal of power-line interference and baseline wander from ECG signals by employing Fourier decomposition technique. *Biomed. Signal Process. Control* 57:101741. doi: 10.1016/j.bspc.2019.101741

Sun, Y., Wei, C., Cui, V., Xiu, M., and Wu, A. (2020). Electroencephalography: clinical applications during the perioperative period. *Front. Med.* 7:251. doi: 10.3389/fmed.2020.00251

Taran, S., Bajaj, V., Sharma, D., Siuly, S., and Segur, A. (2018). Features based on analytic IMF for classifying motor imagery EEG signals in BCI applications. *Measurement* 116, 68–76. doi: 10.1016/j.measurement.2017.10.067

Wang, C., Kipping, J., Bao, C., Ji, H., and Qiu, A. (2016). Cerebellar functional parcellation using sparse dictionary learning clustering. *Front. Neurosci.* 10:188. doi: 10.3389/fnins.2016.00188

Wu, Z. H., and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.* 460, 1597–1611. doi: 10.1098/rspa.2003.1221

Yang, G. A., Yang, J. J., Lu, Z. Z., and Liu, D. Y. (2020). A convolutional neural network with sparse representation. *Knowl. Based Syst.* 209:106419. doi: 10.1016/j.knosys.2020.106419

Yang, Z., Zhuang, X., Bird, C., Sreenivasan, K., Mishra, V., Banks, S., et al. (2019). Performing sparse regularization and dimension reduction simultaneously in multimodal data fusion. *Front. Neurosci.* 13:642. doi: 10.3389/fnins.2019.00642

Zhang, Z. Q., Yang, Q. Y., and Zi, Y. Y. (2021). Multi-scale and multi-pooling sparse filtering: a simple and effective representation learning method for intelligent fault diagnosis. *Neurocomputing* 451, 138–151. doi: 10.1016/j.neucom.2021.04.066

Conflict of Interest: J-LT, Z-FL, and RZ were employed by the company Shaanxi Aerospace Technology Application Research Institute 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.

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 November 2021
doi: 10.3389/fnins.2021.779231

Selective Cross-Subject Transfer Learning Based on Riemannian Tangent Space for Motor Imagery Brain-Computer Interface

Yilu Xu[1], Xin Huang[2] and Quan Lan[3]**

[1] School of Software, Jiangxi Agricultural University, Nanchang, China, [2] Software College, Jiangxi Normal University, Nanchang, China, [3] Department of Neurology, First Affiliated Hospital of Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:
Yizhang Jiang, Jiangnan University, China

Reviewed by:
Shipu Xu, Agricultural Information Institute of Science and Technology, Shanghai Academy of Agricultural Sciences, China
Yuanqing Xian, Guangdong Ocean University, China
Wei Zhou, Shenyang Aerospace University, China

***Correspondence:**
Xin Huang
xinhuang@jxnu.edu.cn
Quan Lan
xmdylanquan@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *18 September 2021*
Accepted: *14 October 2021*
Published: *03 November 2021*

Citation:
Xu Y, Huang X and Lan Q (2021) Selective Cross-Subject Transfer Learning Based on Riemannian Tangent Space for Motor Imagery Brain-Computer Interface. Front. Neurosci. 15:779231. doi: 10.3389/fnins.2021.779231

A motor imagery (MI) brain-computer interface (BCI) plays an important role in the neurological rehabilitation training for stroke patients. Electroencephalogram (EEG)-based MI BCI has high temporal resolution, which is convenient for real-time BCI control. Therefore, we focus on EEG-based MI BCI in this paper. The identification of MI EEG signals is always quite challenging. Due to high inter-session/subject variability, each subject should spend long and tedious calibration time in collecting amounts of labeled samples for a subject-specific model. To cope with this problem, we present a supervised selective cross-subject transfer learning (sSCSTL) approach which simultaneously makes use of the labeled samples from target and source subjects based on Riemannian tangent space. Since the covariance matrices representing the multi-channel EEG signals belong to the smooth Riemannian manifold, we perform the Riemannian alignment to make the covariance matrices from different subjects close to each other. Then, all aligned covariance matrices are converted into the Riemannian tangent space features to train a classifier in the Euclidean space. To investigate the role of unlabeled samples, we further propose semi-supervised and unsupervised versions which utilize the total samples and unlabeled samples from target subject, respectively. Sequential forward floating search (SFFS) method is executed for source selection. All our proposed algorithms transfer the labeled samples from most suitable source subjects into the feature space of target subject. Experimental results on two publicly available MI datasets demonstrated that our algorithms outperformed several state-of-the-art algorithms using small number of the labeled samples from target subject, especially for good target subjects.

Keywords: transfer learning, cross-subject, source selection, Riemannian tangent space, motor imagery

INTRODUCTION

A motor imagery (MI) brain-computer interface (BCI) has drawn great attention for decades, since it can help a subject directly manipulate an electronic equipment using his brain activity evoked by imagined movements, without the participation of the traditional muscle-dependent pathway (He et al., 2015). It can not only help the stroke patients recover their neurological disorders, but also

give able-bodied people a novel way to control an external device (Wolpaw et al., 2002). Therefore, it plays an important role in rehabilitation engineering, military, and entertainment, etc.

Non-invasive electroencephalogram (EEG) is a popular recording modality in MI BCI due to its high safety and high temporal resolution, which is extremely crucial for the application of real-time BCI. However, MI EEG data analysis is quite challenging. The reason is that EEG signals are inherently weak, non-stationary, and easily contaminated by interference and noise (Sample et al., 2019). Moreover, compared with other traditional EEGs, such as event-related potentials (ERP) and steady-state visual evoked potentials (SSVEP), MI EEG signals have less obvious features, because they are evoked by spontaneous movement imagination without external stimulus, whereas ERP and SSVEP are invoked by some external stimulation (Blankertz et al., 2011; Yin et al., 2015). Consequently, MI EEG signals have higher inter-session/subject variability and fewer categories of BCI tasks than ERP and SSVEP. In MI BCI, each subject needs a tedious and annoying calibration time for a subject-specific classifier before performing real-time BCI tasks. A retraining session always increases user frustration.

Deep learning is a promising machine learning technique, which has been widely used in natural language processing and computer vision and so on Huang et al. (2019, 2020). However, it cannot be directly applied to the small training set scenario.

Transfer learning (TL) and semi-supervised learning (SSL) have been tried to reduce the need of abundant labeled samples. TL transfers the labeled samples from different source domains into the target domain. A domain denotes a subject, a session, a task, or a device. SSL simultaneously utilizes the labeled and unlabeled samples from the same subject. Both TL and SSL can shorten the calibration effort for the target subject by utilizing the samples as much as possible.

Here, we pay more attention to TL since it can use more samples than SSL. Generally, TL can be categorized into three groups, including instance TL (ITL), feature-representation transfer (FRT), and parameter TL (PTL) (Pan and Yang, 2010). Instance TL aims to transfer parts of the data in the source domains by reweighting. FRT tries to transfer a good feature representation for the target domain. PTL successively transfers and updates some parameters of the model under the assumption that the source and target domains share these parameters.

Among these groups, ITL approaches are most popular since they are easy to implement. Specifically, they can be divided into two categories.

The first category is filter based ITL, which aims to reweight the labeled sets from different domains based on the filters. In MI BCI, the common spatial patterns (CSP) approach is only effective when the labeled samples from a subject are abundant (Ramoser et al., 2000). Recently, regularized CSP (RCSP) approaches, belonging to filter based ITL, have been designed for the small training set scenario. In the framework of RCSP, the filtered EEG samples from the source and target subjects are separately reweighted based on their similarities which are always measured using different metrics, such as kullback-leibler (KL) divergence, Frobenius norm, cosine distance and so on (Kang et al., 2009; Cheng et al., 2017; Xu et al., 2019). A RCSP based on dynamic time warping (DTW-RCSP) approach aligned the labeled samples from all source subjects to the average of a few target samples from the same class (Azab et al., 2020). In the module of feature extraction, RCSP approaches can generate more reliable CSP spatial filters for the target subject by effectively utilizing the labeled samples from the source subjects. However, calculating the optimal regularization terms might impose computational burden.

The second category is data alignment based ITL. On past decade, Riemannian alignment- (RA-) based ITL approaches have drawn a growing attention in EEG-based BCI since affine transformation can make the covariance matrices from different domains similar. Zanini et al. (2018) proposed a RA-based ITL approach to center the covariance matrices from each domain with respect to their reference matrix. Such reference matrix is the Riemannian mean of the covariance matrices of some resting trials in the corresponding domain. All re-centered covariance matrices from different source domains are concatenated altogether to train a minimum distance to mean (MDM) classifier based on Riemannian Gaussian distributions. Rodrigues et al. (2019) proposed a Riemannian procrustes analysis (RPA) approach using the steps of re-centering, stretching, and rotation for the covariance matrices from different domains to minimize the differences between the source and target domains as much as possible. Zhang and Wu (2020) presented a cross-subject manifold embedded knowledge transfer (MEKT) approach to boost zero-training for the target subject by combining the labeled samples from the source subjects with the unlabeled samples from the target subject. MEKT first performs RA to generate aligned covariance matrices for all subjects and then converts all aligned matrices into the tangent space feature vectors. Finally, MEKT finds optimal projection matrices for the tangent space feature vectors from the source and target subjects to reduce the joint probability distribution shift between subjects. Since most classifiers are designed for the Euclidean space instead of the Riemannian manifold, Euclidean alignment- (EA-) based approaches extend RA-based approaches in the Euclidean space using the Euclidean mean as the reference matrix (He and Wu, 2020). Both RA and EA can shorten the differences between the covariance matrices from the source and target subjects.

Inspired by MEKT, we focus on cross-subject TL based on Riemannian tangent space. In the calibration phase, labeled samples from target subject are successively processed by the modules of signal preprocessing, feature extraction, and classification to build a subject-specific classifier. To reduce the calibration time, the labeled samples from different source subjects can be transferred into the same modules for a more robust classifier.

Nevertheless, most cross-subject TL approaches utilize the labeled samples from all source subjects. Lotte and Guan (2011) designed a RCSP framework which evaluated the differences between the whole CSP feature set from all source subjects and the CSP feature set from the target subject, and then added two different weights for these two sets. Zhang and Wu (2020) found two optimal projection matrices for the whole tangent space vector set from all source subjects and the unlabeled tangent

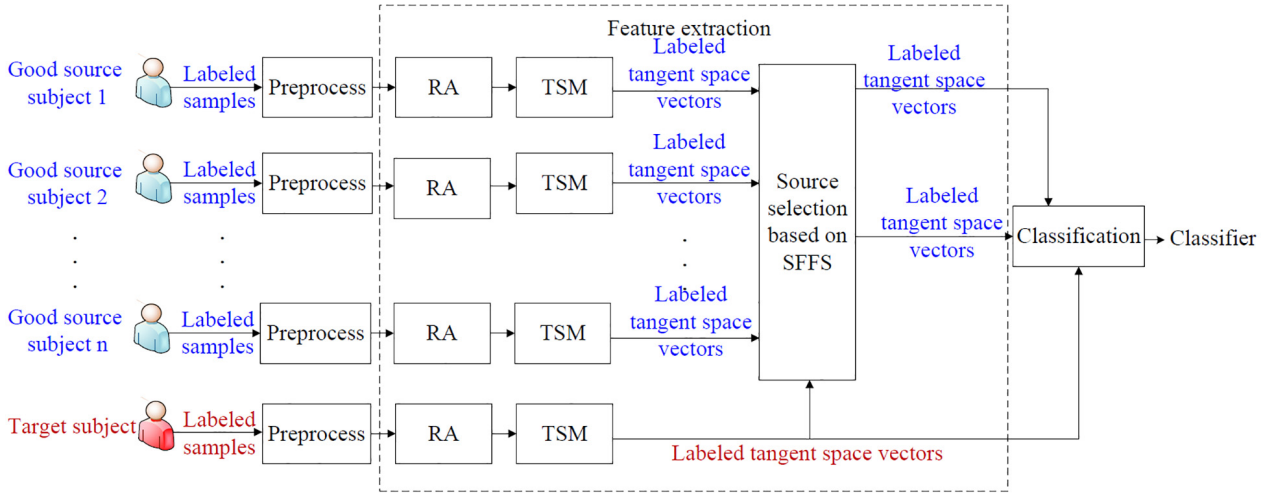


FIGURE 1 | The framework of supervised selective cross-subject transfer learning (sSCSTL).

space vector set from the target subject to lower their dimensions and differences.

To promote the positive transfer, the appropriate source subjects should be selected. Zanini et al. (2018) transferred all good source subjects (those with a precision higher than 0.75) for the target subject. To select suitable source subjects, Liang and Ma (2020) temporarily assigned the labeled samples from the target subject as the testing set, and the labeled samples from the candidate source subject as the training set. Then, the source subjects with higher classification accuracy were selected. Zhang and Wu (2020) presented a domain transferability estimation (DTE) approach based on MEKT, which selected the source subjects with the highest transferability.

In this paper, we first propose a supervised selective cross-subject TL (supervised SCSTL, sSCSTL) approach based on Riemannian tangent space for MI classification, which collects only a few labeled samples from target subject and transfers the labeled samples from most suitable source subjects. Moreover, we present an unsupervised SCSTL (uSCSTL) approach and a semi-supervised SCSTL (ssSCSTL) approach to investigate the role of the unlabeled samples from target subject. A sequential forward floating search (SFFS) method (Pudil et al., 1994) is used to iteratively select suitable source subjects for all SCSTL approaches.

The contribution of this paper lies in two aspects. Firstly, we iteratively select appropriate source subjects, instead of simply selecting them based on their classification performance, transferability and so on. Secondly, we consider SCSTL using different versions to exploit the labeled and unlabeled samples from target subject.

The remainder of this paper is structured as follows. In Section "Methods," our SCSTL approaches are described in detail. In Section "Experimental Results," the effectiveness of our approaches is validated by some experimental results. A discussion of results is presented in Section "Discussion." Finally, our conclusions are drawn in Section "Conclusion."

METHODS

We propose a supervised selective cross-subject transfer learning (sSCSTL) approach which utilizes the labeled samples from the selected good source subjects and the target subject simultaneously. In our opinion, the source subjects with good classification performance are more eligible to be selected than those with bad performance, since the labeled samples from the bad source subjects might mess up the feature distribution of the target subject. The framework of sSCSTL is shown in **Figure 1**.

In **Figure 1**, the labeled samples from each subject are band-pass filtered in the signal preprocessing module. During the feature extraction phase, the filtered labeled samples from each subject are first converted into the labeled covariance matrices. To preliminarily shorten the differences between subjects, the labeled covariance matrices from each subject are transformed into the aligned labeled covariance matrices by performing RA. To train a classifier in the Euclidean space, the aligned labeled matrices from each subject are converted into corresponding labeled tangent space vectors by executing tangent space mapping (TSM). To obtain best source selection, based on the SFFS method, the labeled tangent space vectors from different good source subjects are selected into the selected tangent space vector set by comparing with the labeled tangent space vector set from the target subject. Finally, the labeled tangent space vectors from the most suitable good source subjects and the target subject are fed into the classification module to obtain a subject-specific classifier.

Next, we introduce the sSCSTL approach in detail.

Riemannian Alignment

In MI BCI, supervised CSP is the most popular feature extraction algorithm for single user. Covariance matrices from different classes are used to learn optimal spatial filters which can maximize the variance differences between two classes. Such

spatial filters are used to extract low-dimensional features for the labeled and unlabeled samples. If the labeled samples are noisy, and/or few, covariance matrices might generate unreliable spatial filters. In this fashion, covariance matrices are processed in the Euclidean space. However, they belong to a smooth Riemannian manifold of symmetric positive definite (SPD) matrices, instead of Euclidean space. Thus, in this paper, we first handle covariance matrices in the differentiable Riemannian manifold.

All SPD matrices can form a Riemannian manifold. Here are the basic concepts of SPD matrix and its manifold. Let P_i and P_j be the ith SPD matrix and the jth SPD matrix, respectively, where $P_i, P_j \in \mathbb{R}^{N_c \times N_c}$ and N_c is the number of channels. Their manifold is $N_c \times (N_c + 1)/2$ dimensional. P_i and P_j can be regarded as the points of the manifold. The Riemannian distance $\delta(P_i, P_j)$ between P_i and P_j is the length of the minimum curve connecting them. It can be calculated by Moakher (2005).

$$\delta\left(P_i, P_j\right) = ||\log({P_i}^{-1}P_j)||_{\mathrm{F}} = \left[\sum\nolimits_{k=1}^{N_c} \log^2\lambda_k\right]^{1/2} \tag{1}$$

where $||\cdot||_{\mathrm{F}}$ denotes the Frobenius norm, $\{\lambda_k\}_{k=1}^{N_c}$ are the eigenvalues of ${P_i}^{-1}P_j$. There exist important properties for the Riemannian distance as follows:

$$\delta\left(P_i, P_j\right) = \delta\left(P_j, P_i\right)$$

$$\delta\left(P_i, P_j\right) = \delta\left({P_i}^{-1}, {P_j}^{-1}\right)$$

$$\delta\left(P_i, P_j\right) = \delta\left(W^{\mathrm{T}}P_iW, W^{\mathrm{T}}P_jW\right), \forall W \in \mathrm{Gl}(N_c)$$

The third property is crucial for the context of signal processing. It is named congruence invariance which means that the distance between the two SPD matrices is invariant after affine transformation using an invertible matrix. Suppose that $\mathrm{Gl}(N_c)$ is the set of all $N_c \times N_c$ invertible matrices belonging to the space of square real matrices. The center of all SPD matrices is generally used for affine transformation, which can be calculated using Riemannian distance.

The Riemannian mean M_R, namely geometric mean, is the center point of the manifold which can be calculated using Riemannian distance as follows:

$$M_R = \underset{M}{\operatorname{argmin}} \sum\nolimits_{\mathrm{k}=1}^{N} \delta^2(P_k, M), \tag{2}$$

where N denotes the number of all SPD matrices on the manifold. There is no closed-form way to calculate the Riemannian mean. An iterative method can be used to effectively obtain the Riemannian mean (Fletcher and Joshi, 2004).

We can perform affine transformation using the center of all SPD matrices as the reference matrix. RA executes the following transformation using the Riemannian mean M_R as the reference matrix:

$$\acute{P}_i = {M_R}^{-1/2}P_i{M_R}^{-1/2}. \tag{3}$$

Likewise, EA extends RA using the Euclidean mean M_E as the reference matrix (He and Wu, 2020). The Euclidean mean M_E, namely arithmetic mean, is defined as the center point which can minimize the sum of squared Euclidean distances between other matrices and the center point as below:

$$M_E = \underset{M}{\operatorname{argmin}} \sum\nolimits_{\mathrm{k}=1}^{N} d^2\left(P_k, M\right) = \frac{1}{N}\sum\nolimits_{\mathrm{k}=1}^{N} P_k. \tag{4}$$

where $d(\cdot, \cdot)$ computes the Euclidean distance between the two SPD matrices. Since the Euclidean mean can be easily computed, it is widely used so far.

In this paper, we perform RA for all covariance matrices from each subject since RA is more suitable for the Riemannian manifold.

For RA-based ITL approaches, all aligned covariance matrices from each domain are centered at the identity matrix after affine transformation using each domain's Riemannian mean as the reference matrix. This property makes the aligned covariance matrices from different domains comparable, which can be testified by the following (Zhang and Wu, 2020).

$$\begin{aligned}&\mathcal{M}\left({M_R}^{-1/2}P_1{M_R}^{-1/2}, {M_R}^{-1/2}P_2{M_R}^{-1/2}\right.\\&\left., \cdots, {M_R}^{-1/2}P_N{M_R}^{-1/2}\right)\\&= {M_R}^{-1/2}\mathcal{M}(P_1, P_2, \cdots, P_N){M_R}^{-1/2}\\&= {M_R}^{-1/2}M_R{M_R}^{-1/2} = I\end{aligned} \tag{5}$$

where $\mathcal{M}(P_1, P_2, \cdots, P_N)$ is the Riemannian mean operation, and I is an identity matrix.

Moreover, each aligned covariance matrix is nearly whitened since it is approximately an identity matrix. More details can be seen in Zhang and Wu (2020). Consequently, RA can preliminarily reduce the inter-domain differences.

Tangent Space Mapping

After RA, the aligned covariance matrices are usually input to an MDM classifier (Zanini et al., 2018; Rodrigues et al., 2019). However, most classifiers, such as linear discriminant analysis (LDA) and support vector machine (SVM), are designed for the Euclidean space. Thus, we perform tangent space mapping (TSM) for all aligned matrices from each subject.

As mentioned above, all SPD matrices lie in a differentiable Riemannian manifold. Then their derivatives at a matrix on the manifold form a tangent space. The tangent space has the same dimensions as the manifold. A Riemannian manifold and its tangent space at a point can be illustrated in **Figure 2**.

As depicted in **Figure 2**, P_i and P_j belong to the Riemannian manifold. On the tangent space at point M, A_i and A_j are the derivatives of P_i and P_j, respectively. A_i can be regarded as the logarithmic mapping of P_i at M as below:

$$A_i = \log_M(P_i) = M^{1/2}\log(M^{-1/2}P_iM^{-1/2})M^{1/2}. \tag{6}$$

Conversely, the exponential mapping of A_i at M is the SPD matrix P_i defined as:

$$P_i = \exp_M(A_i) = M^{1/2}\exp(M^{-1/2}\mathrm{A}_iM^{-1/2})M^{1/2}. \tag{7}$$

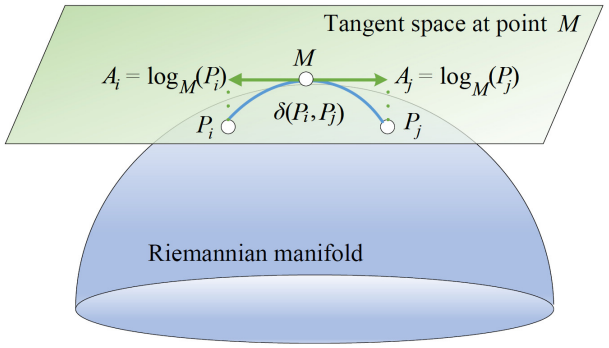


FIGURE 2 | A Riemannian manifold and its tangent space at a point.

Riemannian distance can be also defined as:

$$\begin{aligned}\delta(M, P_i) &= ||\log_M(P_i)||_M = ||A_i||_M \\ &= ||\text{vec}(M^{-1/2}A_iM^{-1/2})||_2 = ||\text{vec}(\log(M^{-1/2}P_iM^{-1/2}))||_2 \\ &= ||\text{vec}(\acute{A}_i)||_2 = ||\text{a}_i||_2, \end{aligned} \tag{8}$$

where $||A_i||_M$ is the norm of A_i on the tangent space at point M, and $\text{vec}(\acute{A}_i)$ vectorizes a symmetry matrix $\acute{A}_i$. Let $\acute{A}_i$ and a_i be $\log(M^{-1/2}P_iM^{-1/2})$ and $\text{vec}(\acute{A}_i)$, respectively. Without loss of generality, we apply a $\sqrt{2}$ coefficient for out-of-diagonal elements of $\acute{A}_i$ and transform the upper triangular part of modified $\acute{A}_i$ into a $N_c \times (N_c + 1)/2$ column vector a_i as below:

$$\begin{aligned}\text{a}_i = \text{vec}(\acute{A}_i) = [&\acute{A}_{i_{1,1}}; \sqrt{2}\acute{A}_{i_{1,2}}; \acute{A}_{i_{2,2}}; \sqrt{2}\acute{A}_{i_{1,3}}; \sqrt{2}\acute{A}_{i_{2,3}}; \acute{A}_{i_{3,3}} \\ &; \cdots; \acute{A}_{i_{N_c,N_c}}] \end{aligned} \tag{9}$$

where $\acute{A}_{ij,k} \in \acute{A}_i$. Then, a_i is the tangent space vector of the SPD matrix P_i at point M.

Furthermore, there is an approximation in terms of distance between the manifold and its tangent space as below (Tuzel et al., 2008):

$$\delta\left(P_i, P_j\right) \approx ||\text{a}_i - \text{a}_j||_2 \tag{10}$$

where P_i and P_j are locally distributed into the manifold. Their tangent space vectors at point M are a_i and a_j, respectively. Such approximation stands only when M is the Riemannian mean of the manifold.

Therefore, Riemannian tangent space is Euclidean and locally homomorphic to the Riemannian manifold (Barachant et al., 2013). In the TSM phase of our proposed sSCSTL, all aligned covariance matrices $\{{M_R}^{-1/2}P_i{M_R}^{-1/2}\}_{i=1}^N$ from each domain are converted into corresponding tangent space vectors $\{\text{a}_i = \text{vec}(\log({M_R}^{-1/2}P_i{M_R}^{-1/2}))\}_{i=1}^N$ using M_R as the reference matrix.

Source Selection Based on Sequential Forward Floating Search

As mentioned above, due to high inter-subject variability and expensive computational burden, it is unsuitable to utilize the features from all source subjects. In this paper, we only transfer the labeled tangent space vectors from most suitable good source subjects.

We define the transferability for the selected tangent space vector set *Sel* to predict its usefulness for the tangent space vector set $\mathcal{D}_T$ from the target subject as in Zhang and Wu (2020):

$$transferability(Sel, \mathcal{D}_T) = \frac{||{S_b}^{Sel}||_1}{||{S_b}^{Sel,\ \mathcal{D}_T}||_1}, \tag{11}$$

where ${S_b}^{Sel}$ is the between-class scatter matrix of *Sel*. Similarly, ${S_b}^{Sel,\ \mathcal{D}_T}$ is the scatter matrix between *Sel* and $\mathcal{D}_T$. Then, $||{S_b}^{Sel}||_1$ is used for the discriminability of between-class of *Sel*, and $||{S_b}^{Sel,\ \mathcal{D}_T}||_1$ is used to evaluate the differences between the selected source subjects and the target subject.

Instead of simply integrating the labeled tangent space vector sets from all source subjects, the labeled tangent space vector set from the good source subject is iteratively selected to add into or remove from the selected tangent space vector set to maximize its transferability. Our proposed source selection procedure for sSCSTL uses the framework of SFFS to obtain convergence of selection. More details are shown in **Algorithm 1**.

Algorithm 1: The source selection procedure for sSCSTL
Input: N_{TL} labeled tangent space vectors from the target subject $\mathcal{D}_{TL} = \{\text{a}_i^{(TL)}, y_i^{(TL)}\}_{i=1}^{N_{TL}}$, N_{S_k} labeled tangent space vectors from the kth good source subject $\mathcal{D}_{S_k} = \{\text{a}_i^{(S_k)}, y_i^{(S_k)}\}_{i=1}^{N_{S_k}}$, $k = \{1, 2, \cdots, N_{GS}\}$, where N_{GS} is the number of good source subjects. Let $y_i^{(TL)}$ and $y_i^{(S_k)}$ be the labels of tangent space vectors from target and source subjects, respectively.
Output: The finally selected tangent space vector set *Sel*;
Initialize: The initially selected tangent space vector set $Sel_0 = \emptyset$ and its transferability $Trans_0 = 0$, the initially remaining tangent space vector set $Rem_0 = \{\mathcal{D}_{S_k}\}_{k=1}^{N_{GS}}$, $n = 1$;
Repeat
Step 1: The labeled tangent space vector set from the most suitable good source subject is selected and added into Sel_{n-1}

$$\mathcal{D}_{S_{good}} = \underset{\mathcal{D}_{S_k} \in Rem_{n-1}}{\text{argmax}}\ (transferability(Sel_{n-1} + \mathcal{D}_{S_k}, \mathcal{D}_{TL}));$$

$$Sel_n = Sel_{n-1} + \mathcal{D}_{S_{good}};\ Rem_n = Rem_{n-1} - \mathcal{D}_{S_{good}};$$

$$Trans_n = transferability(Sel_n, \mathcal{D}_{TL});$$

$$n = n + 1;$$

Step 2: The labeled tangent space vector set from the most unsuitable good source subject is selected and removed from Sel_n
if $n > 2$ **then**

$$\mathcal{D}_{S_{bad}} = \underset{\mathcal{D}_{S_k} \in Sel_n}{\text{argmax}}(transferability(Sel_n - \mathcal{D}_{S_k}, \mathcal{D}_{TL}));$$

if $transferability(Sel_n - \mathcal{D}_{S_{bad}}, \mathcal{D}_{TL}) > Trans_{n-1}$ **then**

$$Sel_{n-1} = Sel_n - \mathcal{D}_{S_{bad}};\ Rem_{n-1} = Rem_n + \mathcal{D}_{S_{bad}};$$

$Trans_{n-1} = transferability(Sel_{n-1}, \mathcal{D}_{TL})$;

$n = n - 1$;

go to Step 2;

else go to Step 1;
end
end
Until $n = N_{GS}$

$Sel = Sel_{arg\max_k(Trans_k)}$.

For sSCSTL, labeled covariance matrices from each subject are converted into labeled tangent space vectors using their Riemannian mean as the reference matrix.

In **Algorithm 1**, during each iteration, Sel_n and $Trans_n$ separately denote the currently selected tangent space vector set from n good source subjects and its transferability, whereas Rem_n denotes the currently remaining tangent space vector set from $(N_{GS} - n)$ good source subjects. After all iterations, the selected tangent space vector set with the highest transferability is chosen as the finally selected tangent space vector set *Sel*.

Since $||S_b^{Sel,\ \mathcal{D}_T}||_1$ in Equation (11) is evaluated without using the labels of *Sel* and $\mathcal{D}_T$, we modify the source selection procedure in **Algorithm 1** for an unsupervised SCSTL (uSCSTL) approach and a semi-supervised SCSTL (ssSCSTL) approach to investigate the role of unlabeled samples from the target subject.

For uSCSTL, unlabeled covariance matrices from target subject are converted into corresponding tangent space vectors $\mathcal{D}_{TU}$ using their Riemannian mean as the reference matrix. Then, in each iteration, the currently selected tangent space vector set Sel_n is compared with $\mathcal{D}_{TU}$ to obtain its transferability ($transferability(Sel_n,\ D_{TU})$).

For ssSCSTL, all covariance matrices from target subject are projected into corresponding tangent space vectors ($\mathcal{D}_T = \mathcal{D}_{TL} + \mathcal{D}_{TU}$) using their Riemannian mean as the reference point. Then, more tangent space vectors from target subject can be used to yield more convincing transferability ($transferability(Sel_n,\ D_T)$).

Classification

The supervised LDA classifier has been widely used because of simplicity and rapid computational speed, which is crucial for real time MI BCI. Nevertheless, it is of poor quality for high dimensional features, such as the tangent space vectors. Therefore, we use a supervised shrinkage LDA (sLDA) classifier (Peck and Van Ness, 1982) which not only inherits the advantages of LDA but also behaves well in high dimensions.

Our goal of sSCSTL, uSCSTL, and ssSCSTL is to extract discriminative features for the classifier. In our classification module, no unlabeled tangent space vectors from the target subject are utilized for the classifier. Therefore, for sSCSTL and ssSCSTL, the labeled tangent space vectors from the selected good source subjects and target subject are fed into the sLDA classifier after the best source selection based on SFFS, whereas for uSCSTL, only the labeled tangent space vectors from the selected good source subjects are used for sLDA.

EXPERIMENTAL RESULTS

Datasets

In this paper, the MI dataset from 52 healthy subjects (Cho et al., 2017) and the BCI competition IV dataset 1 (Tangermann et al., 2012) were used to evaluate the effectiveness of our proposed methods in MI classification.

(1) MI dataset from 52 healthy subjects (MI1): 64-channel MI EEG signals from 52 healthy subjects were recorded at a sampling rate of 512 Hz. Each subject executed the assigned MI task when a random instruction ("left hand" or "right hand") appeared on the screen for 3 s. Only a single session was performed for each subject, which consisted of five or six runs. In each run, the MI experiments per class were repeated 20 times. Consequently, there were a total of 200 or 240 trials per subject. This dataset was much suitable for cross-subject TL due to the existence of many subjects.
(2) BCI competition IV dataset 1 (MI2): this dataset was comprised of the calibration data and the evaluation data. However, we only chose the calibration data for our experiments since the evaluation data contained the periods in which the subject had no control intention. For the calibration data, 59-channel EEG signals from seven healthy subjects (a, b, c, d, e, f, and g) were collected at a sampling rate of 100 Hz. Each subject was shown a visual cue for 4 s and performed the cued MI task ("left hand" or "foot"). In total, for each subject, 100 trials per class were gathered. All EEG signals were band-pass filtered between 0.05 Hz and 200 Hz.

Electroencephalogram Data Preprocessing

For dataset MI1, the BBCI toolbox (Blankertz et al., 2016) was used for preprocessing. All EEG signals from each subject were preprocessed by common average reference, spectrally filtered by a third order Butterworth filter with cutoff frequencies of 8 and 30 Hz, and temporally segmented from 0.5 to 2.5 s after the instruction onsets.

For dataset MI2, all EEG recordings from each subject were band-pass filtered between 8 and 30 Hz using a fiftieth order finite impulse response filter. Then, the filtered EEG recordings were extracted from the time interval between 0.5 and 3.5 s after the visual cue onsets.

As mentioned above, the dimensionality of tangent space vectors closely relates to the number of channels. High-dimensional tangent space vectors may lead to curse of dimensionality. Therefore, for datasets MI1 and MI2, 27 and 29 channels were separately selected from the sensorimotor areas. In **Figure 3**, the selected channels used for datasets MI1 and MI2 are marked in green.

Experimental Design

First, we selected the good subjects as the source subjects based on their high classification accuracies. In our experiments, the labels of their samples were assumed to be known already. All their samples were used for cross-subject TL. If one of the good

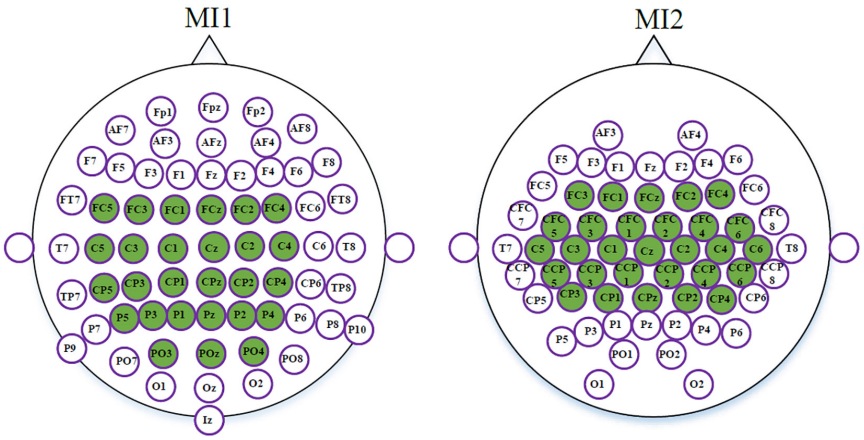


FIGURE 3 | The selected channels for MI1 and MI2.

subjects was the target subject, the remaining good subjects were his source subjects. All samples from each target subject were randomly partitioned into the labeled set for training and the unlabeled set for testing. To avoid randomness involved, this process was repeated ten times and the average accuracies of the unlabeled set were reported. The ratio of number of labeled samples to all samples (R_l) ranged from 10 to 40% with step 10%. Correspondingly, the ratio of number of unlabeled samples to all samples (R_u) varied from 90 to 60%.

Baseline Algorithms

We compared our proposed algorithms (sSCSTL, ssSCSTL, and uSCSTL) with the following baseline algorithms. They can be grouped into three categories according to their feature spaces:

(1) Euclidean space algorithms:

   (a) CSP (Ramoser et al., 2000): it was a traditional feature extraction algorithm for MI. We used LDA as its classifier.
   (b) CSP with generic learning regularization (GLRCSP) (Lu et al., 2009): it learned generic learning regularized CSP filters to realize cross-subject TL. LDA was its classifier.
   (c) Composite CSP (CCSP) (Kang et al., 2009): it weighed covariance matrices from different subjects using the KL divergence measure based on the framework of RCSP. LDA was also its classifier.

(2) Riemannian space algorithm: all covariance matrices belong to the Riemannian manifold. We designed a covariance matrix -based feature extraction algorithm, named COV, which directly input all labeled covariance matrices into the MDM classifier without RA.
(3) Riemannian Tangent space algorithms:

   (a) Tangent space (TS) feature extraction algorithm (Barachant et al., 2012): it successively performed RA and TSM for all covariance matrices. Then, the tangent space vectors were fed into the LDA classifier.
   (b) Manifold embedded knowledge transfer (MEKT) (Zhang and Wu, 2020): it combined the labeled tangent space vectors from the source subjects with the unlabeled tangent space vectors from the target subject. sLDA was used as its classifier.
   (c) Domain transferability estimation (DTE) (Zhang and Wu, 2020): it identified one most suitable source subject for the target subject based on MEKT. sLDA was also used for MI classification.

We set the hyper-parameters of all baseline algorithms instructed by their publications. Our proposed algorithms and seven baselines can also be further divided into two groups. The first group consists of CSP, GLRCSP, CCSP, COV, TS, sSCSTL, and ssSCSTL. They are all supervised or semi-supervised feature extraction algorithms. The second group includes MEKT, DTE, and uSCSTL. They are unsupervised feature extraction algorithms.

Classification Accuracies

Classification Performance on Dataset MI1

For a small percentage of subjects, their classification accuracies hardly reach the benchmark of 70%. This phenomenon has been named BCI illiteracy. It was suggested that at least 40 labeled samples per class can reduce BCI illiteracy (Blankertz et al., 2007). Therefore, for dataset MI1, we chose the good subjects whose accuracies were over 80% on average when 40 or 48 labeled samples per class (R_l = 40%) were randomly selected for 10 times and trained by CSP and LDA.

In our experiments, we focused on the good and bad subjects to discuss how they influenced the BCI performance. Furthermore, we cared about the supervised, semi-supervised, and unsupervised algorithms to investigate the role of the labeled and unlabeled samples.

The classification performances of supervised and semi-supervised algorithms for good subjects on dataset MI1 are

given in **Table 1** using their 40 or 48 labeled samples per class (R_l = 40%). Then, the performances of unsupervised algorithms for the same good subjects are shown in **Table 2** using their remaining unlabeled samples (R_u = 60%). Meanwhile, the means and standard deviations (Std) of good subjects for two groups of algorithms are listed in **Tables 1**, **2**, respectively.

As shown in **Table 1**, there were only seven good subjects selected from 52 subjects on dataset MI1 according to the performance of CSP. For all Euclidean space-based algorithms, there were no obvious differences among them. CSP slightly outperformed GLRCSP and lagged behind CCSP a little. Without RA, the Riemannian space-based algorithm COV performed the worst. For Riemannian tangent space-based algorithms, the mean of TS was lower than that of CSP. sSCSTL showed its superiority among all supervised algorithms. A paired t-test showed that the result of sSCSTL was statistically higher than that of CSP (p = 0.0067), GLRCSP (p = 0.0055), CCSP (p = 0.0064), COV (p = 0.0052), and TS (p = 0.0077). ssSCSTL stood out itself with the help of unlabeled samples. In **Table 2**, all unsupervised algorithms utilized the labeled samples from the selected source subjects and added the remaining unlabeled samples from the good target subject. MEKT surpassed most of supervised algorithms in **Table 1**, except for sSCSTL. Although DTE had the same framework of MEKT, it performed worse than MEKT. The possible reason was that there was only one suitable source subject selected. For uSCSTL, the average classification accuracies of six good subjects were up to 90%.

To further evaluate the effect of the labeled and unlabeled samples from the target subject, we plotted the means of accuracies of all good subjects on dataset MI1 for different algorithms with varying R_l or R_u in **Figure 4**.

As depicted in **Figure 4**, all supervised and semi-supervised algorithms provided good classification performances with varying R_l from 10 to 40%. Except for COV, most algorithms outperformed CSP when few labeled samples from target subject were available. Moreover, the classification performances of most algorithms improved obviously along with the increasement of R_l. However, sSCSTL and ssSCSTL steadily showed compelling validity. In addition, the curves of unsupervised algorithms improved slowly with the increasement of R_u, suggesting that the labeled samples were more useful for the performance than the unlabeled samples. Nevertheless, even using 60% unlabeled samples from the target subject, the average classification accuracy was over 95% for uSCSTL.

Then, since there were 45 bad subjects on dataset MI1, we did not provide their average classification accuracies one by one. Instead, in **Tables 3**, **4**, we separately investigated how many bad subjects were trapped in BCI illiteracy with varying R_l and R_u. Here, those whose average classification accuracies were less than 70% were regarded as unqualified subjects.

As shown in **Table 3**, Euclidean space-based algorithms sharply reduced the number of unqualified subjects when their labeled samples increased. It implied that they performed well when abundant labeled samples from target subject were available. In contrast, other algorithms reduced BCI illiteracy slowly with the increasement of R_l, suggesting that their performances did not excessively rely on the number of labeled samples. sSCSTL reduced more unqualified subjects than CCSP when R_l was 10%. In **Table 4**, the number of unqualified subjects barely changed with the increasement of R_u for all unsupervised algorithms. uSCSTL also reduced more unqualified subjects than MEKT when R_u was 60%.

Like **Figure 4**, the means of accuracies of all bad subjects on dataset MI1 for different algorithms with varying R_l or R_u are plotted in **Figure 5**.

As illustrated in **Figures 4**, **5**, the advantages of sSCSTL and ssSCSTL for all bad subjects were less apparent than those for good ones. In **Figure 5**, there were few differences between sSCSTL and ssSCSTL. Possible explanation was that low class-discriminability of unlabeled samples from bad target subjects influenced the performance improvements of ssSCSTL. In addition, CCSP and GLRCSP still outperformed CSP, TS, and COV. Additionally, the average classification accuracies provided by uSCSTL were slightly lower than those of MEKT with varying R_u. The performance of DTE showed the importance of using source subjects as much as possible.

TABLE 1 | Average classification accuracies of supervised and semi-supervised algorithms for good subjects on dataset MI1 (R_l = 40%).

	3	4	14	23	41	43	48	Mean (Std)
CSP	89.50	82.50	96.92	81.33	83.67	96.92	82.33	87.60 (6.90)
GLRCSP	89.58	82.92	96.17	80.17	85.83	95.92	80.58	87.31 (6.78)
CCSP	90.08	84.33	96.92	80.75	85.50	96.08	82.08	87.96 (6.54)
COV	86.42	77.33	96.92	73.08	73.92	93.17	75.58	82.35 (9.78)
TS	90.17	77.92	96.25	80.83	80.50	96.08	82.83	86.37 (7.70)
sSCSTL	92.00	95.75	100.00	94.25	97.67	99.83	97.75	96.75 (2.94)
ssSCSTL	91.33	99.42	100.00	99.42	99.42	100.00	100.00	98.51 (3.18)

TABLE 2 | Average classification accuracies of unsupervised algorithms for good subjects on dataset MI1 (R_u = 60%).

	3	4	14	23	41	43	48	Mean (Std)
MEKT	90.42	84.17	95.25	82.17	86.92	94.83	82.75	88.07 (5.51)
DTE	89.42	80.75	95.00	78.58	82.42	92.33	81.92	85.77 (6.38)
uSCSTL	88.83	97.17	99.50	93.83	97.50	100.00	99.17	96.57 (3.99)

Classification Performance on Dataset MI2

On dataset MI2, five good subjects (a, d, e, f, and g) were chosen out of the seven subjects depending on the same rule as mentioned above. Then subjects b and c were regarded as bad subjects. The classification performances of supervised and semi-supervised algorithms for all subjects on dataset MI2 are shown in **Table 5** using their 40 labeled samples per class (R_l = 40%). And the performances of unsupervised algorithms for the same subjects are given in **Table 6** using their remaining unlabeled samples (R_u = 60%).

As shown in **Table 5**, among all Euclidean space-based algorithms, GLRCSP and CCSP outperformed CSP on average. COV still performed the worst due to without RA. For Riemannian tangent space-based algorithms, the average accuracy of TS was slightly lower than that of CSP. sSCSTL

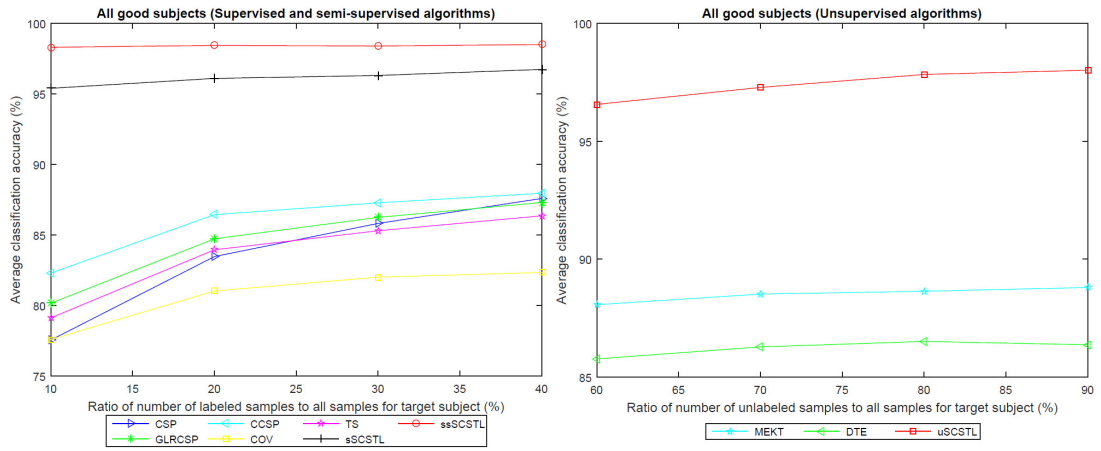


FIGURE 4 | The means of accuracies of all good subjects on dataset MI1 for different algorithms with varying R_l or R_u.

and ssSCSTL stood out themselves on average. However, for bad subjects, GLRCSP and CCSP were comparably superior to other algorithms. For most good subjects (d, e, f, and g), sSCSTL and ssSCSTL exhibited their superiorities. Paired t-test results between sSCSTL and other supervised algorithms showed that the performance improvements of sSCSTL over others were statistically significant, such as sSCSTL vs. CSP (p = 0.1795), sSCSTL vs. GLRCSP (p = 0.2901), sSCSTL vs. CCSP (p = 0.3753), sSCSTL vs. COV (p = 0.0260), and sSCSTL vs. TS (p = 0.1631). As reported in **Table 6**, although uSCSTL showed its compelling validity on average, MEKT performed better than uSCSTL for bad subjects.

To further evaluate the role of labeled and unlabeled samples, for different algorithms, we plotted the average classification accuracies of all subjects on dataset MI2, and their means, with varying R_l and R_u in **Figures 6**, **7**, respectively.

In the last subfigure of **Figure 6**, ssSCSTL outperformed other algorithms on average with the help of unlabeled samples from target subject. The gap between sSCSTL and ssSCSTL became small with the increasement of labeled samples from target subject. For Euclidean space-based algorithms, CCSP and GLRCSP still performed better than CSP on average. Riemannian tangent space-based algorithm TS showed higher performance than Riemannian space-based algorithm COV.

Moreover, as depicted in other subfigures of **Figure 6**, sSCSTL and ssSCSTL showed amazing accuracies for most good subjects (d, e, f, and g), suggesting that the available samples from a good subject can provide positive cross-subject TL. Exceptionally, a good subject (a) benefited little from TL algorithms. Additionally, sSCSTL and ssSCSTL performed more unstably than CCSP and GLRCSP for bad subjects (b and c) in total.

As shown in **Figure 7**, the average classification accuracies of uSCSTL were inferior to those of MEKT and DTE for the good subject (a) and bad subjects (b and c). However, uSCSTL stood out itself for other subjects. Therefore, in the last subfigure of **Figure 7**, uSCSTL performed better than other unsupervised algorithms overall.

TABLE 3 | The number of unqualified subjects whose accuracies were less than 70% on dataset MI1 for supervised and semi-supervised algorithms with varying R_l.

	10%	20%	30%	40%	Mean
CSP	45	42	37	32	39
GLRCSP	44	39	36	31	37.5
CCSP	42	37	33	31	35.75
COV	43	43	43	41	42.5
TS	43	42	40	38	40.75
sSCSTL	35	34	33	34	34
ssSCSTL	34	34	34	31	33.25

TABLE 4 | The number of unqualified subjects whose accuracies were less than 70% on dataset MI1 for unsupervised algorithms with varying R_u.

	60%	70%	80%	90%	Mean
MEKT	45	45	45	45	45
DTE	45	45	45	45	45
uSCSTL	38	37	38	37	37.5

Data Visualization

To appreciate the effect of cross-subject TL, in **Figure 8**, the t-SNE method was applied to visualize the feature distributions of the following Riemannian space-based and Riemannian tangent space-based algorithms, including COV, TS, sSCSTL, ssSCSTL, and uSCSTL (Van Der Maaten and Hinton, 2008). In our case, each feature was regarded as a point in a two-dimensional space. For the dataset MI1, the first 40 samples per class from good target subject 23 were labeled for training, the rest of samples were unlabeled for testing. Additionally, all samples from other good subjects were the labeled source samples.

In **Figure 8**, for COV and TS, the discrimination of two labeled target classes marked in red were obscure, resulting in

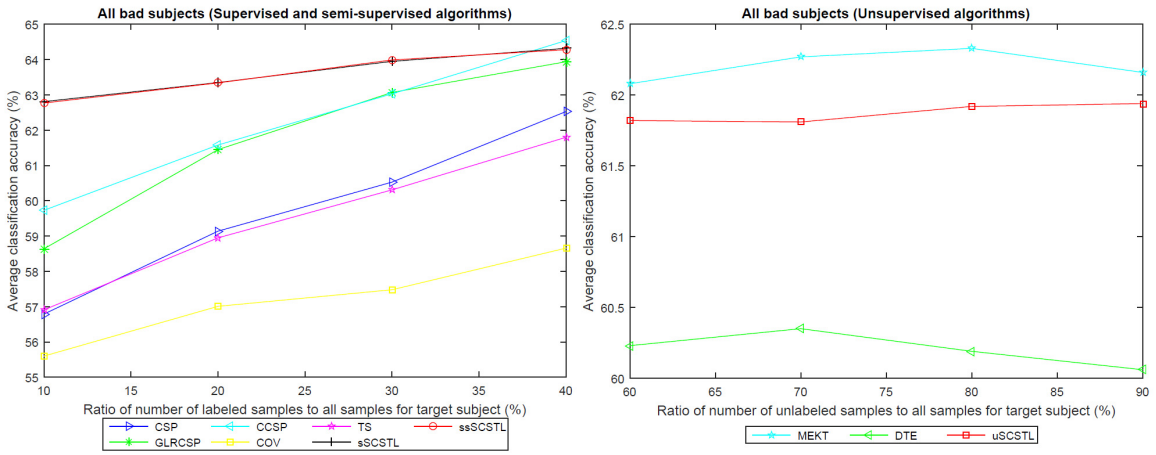


FIGURE 5 | The means of accuracies of all bad subjects on dataset MI1 for different algorithms with varying R_l or R_u.

the difficulty of classification for the unlabeled target samples marked in blue. However, after RA and TSM, the labeled target samples with same class for TS were closer than those for COV. sSCSTL and uSCSTL discriminated the two-class labeled samples from target and source subjects well, after source selection using the labeled and unlabeled target samples, respectively. Moreover, the classification performances of our proposed algorithms might increase since many unlabeled target samples overlapped with labeled source samples which had the same class. Although, ssSCSTL could not separate the two labeled classes well, it performed well in most cases.

Computation Time Comparison

Figure 9 shows average computation time of different algorithms on datasets MI1 and MI2. For supervised and semi-supervised algorithms, 40% of total samples from target subjects are used for training. For unsupervised algorithms, 60% of total samples from target subjects are unlabeled and used for cross-subject TL. Our experimental results were conducted in Matlab 2015a on a laptop with intel i5 CPU@1.60 GHz, 8 GB memory, running 64-bit Windows 10 Home Edition.

As shown in **Figure 9**, GLRCSP and CCSP spend much longer computation time than other algorithms when many good source subjects are available (seven for MI1 and five for MI2). Although CSP, COV, and TS save the running time, SCSTL and ssSCSTL spend acceptable running time. In addition, among all unsupervised algorithms, it takes the lowest time for DTE since DTE only selects one most appropriate source subject. The computation time of uSCSTL is also acceptable, although uSCSTL approximately spends twice running time of MEKT.

TABLE 5 | Average classification accuracies of supervised and semi-supervised algorithms for all subjects on dataset MI2 (R_l = 40%).

	a	b	c	d	e	f	g	Mean (Std)
CSP	87.67	69.33	68.75	87.17	92.50	89.50	91.17	83.73 (10.20)
GLRCSP	86.17	73.33	72.08	84.08	93.75	89.92	89.92	84.18 (8.42)
CCSP	87.67	74.50	71.17	86.75	93.00	90.75	90.92	84.96 (8.60)
COV	81.42	65.50	67.75	76.00	74.50	80.00	76.58	74.54 (5.93)
TS	88.33	65.50	70.83	82.42	89.83	89.33	92.33	82.65 (10.46)
sSCSTL	79.33	69.92	68.08	99.92	100.00	98.17	100.00	87.92 (14.90)
ssSCSTL	83.25	67.50	71.50	100.00	100.00	100.00	100.00	88.89 (14.64)

TABLE 6 | Average classification accuracies of unsupervised algorithms for all subjects on dataset MI2 (R_u = 60%).

	a	b	c	d	e	f	g	Mean (Std)
MEKT	78.33	73.50	73.00	70.00	90.33	80.83	91.92	79.70 (8.59)
DTE	71.00	72.25	71.25	81.08	86.33	68.42	79.33	75.67 (6.61)
uSCSTL	63.00	65.58	62.67	100.00	100.00	99.75	99.92	84.42 (19.35)

DISCUSSION

In this section, we discuss the experimental results from the following aspects.

(1) In terms of cross-subject TL

There are three non-TL algorithms (CSP, COV, and TS) in our paper. As shown in **Figures 4-6**, the average classification accuracies of TL algorithms are superior to those of non-TL algorithms in total. It proves the validity of cross-subject TL methods when few labeled samples from target subject are available, thus effectively shortening the calibration time for target subject. As depicted in **Figure 9**, compared with non-TL algorithms, most cross-subject TL algorithms spend more but acceptable running time except for GLRCSP and CCSP. However, the cross-subject TL algorithms in our paper are all offline algorithms, in which all samples from target and source subjects are ready in advance. Therefore, they are not suitable for real-time BCI applications.

(2) In terms of different feature spaces

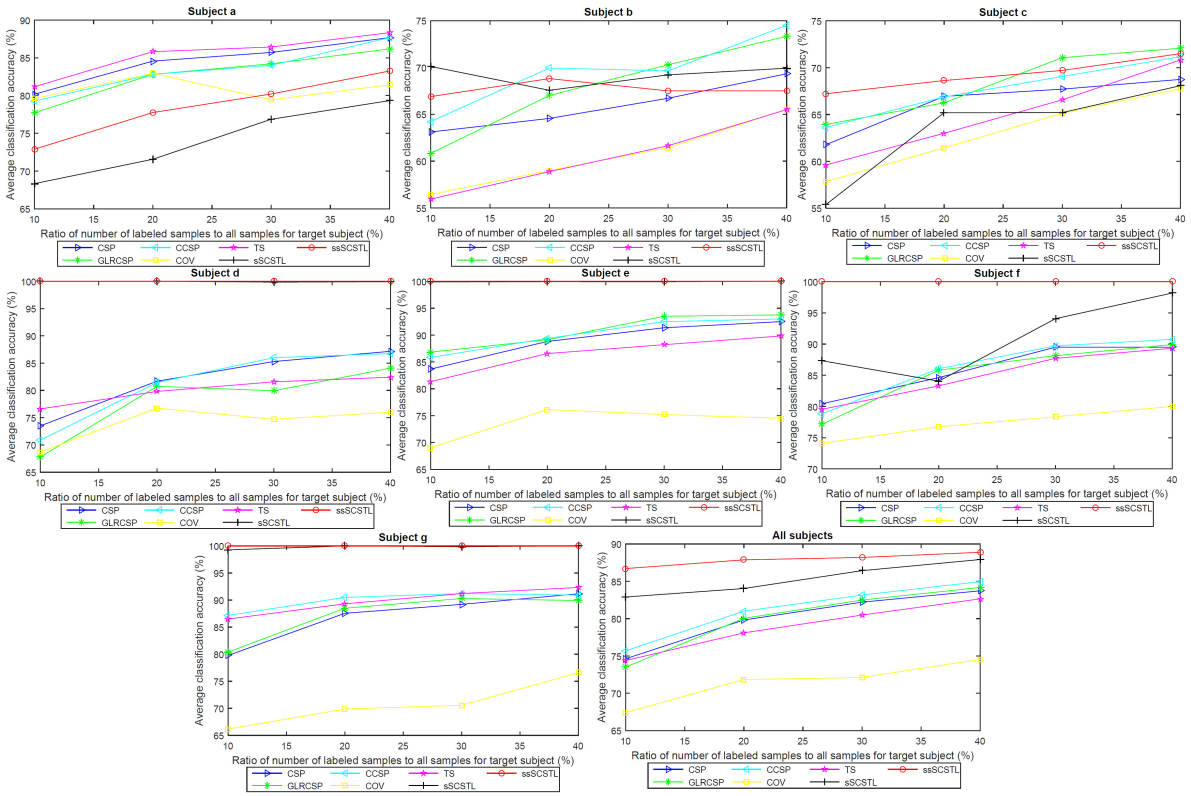


FIGURE 6 | Average classification accuracies of all subjects on dataset MI2, and their means, with varying R_l, for supervised and semi-supervised algorithms.

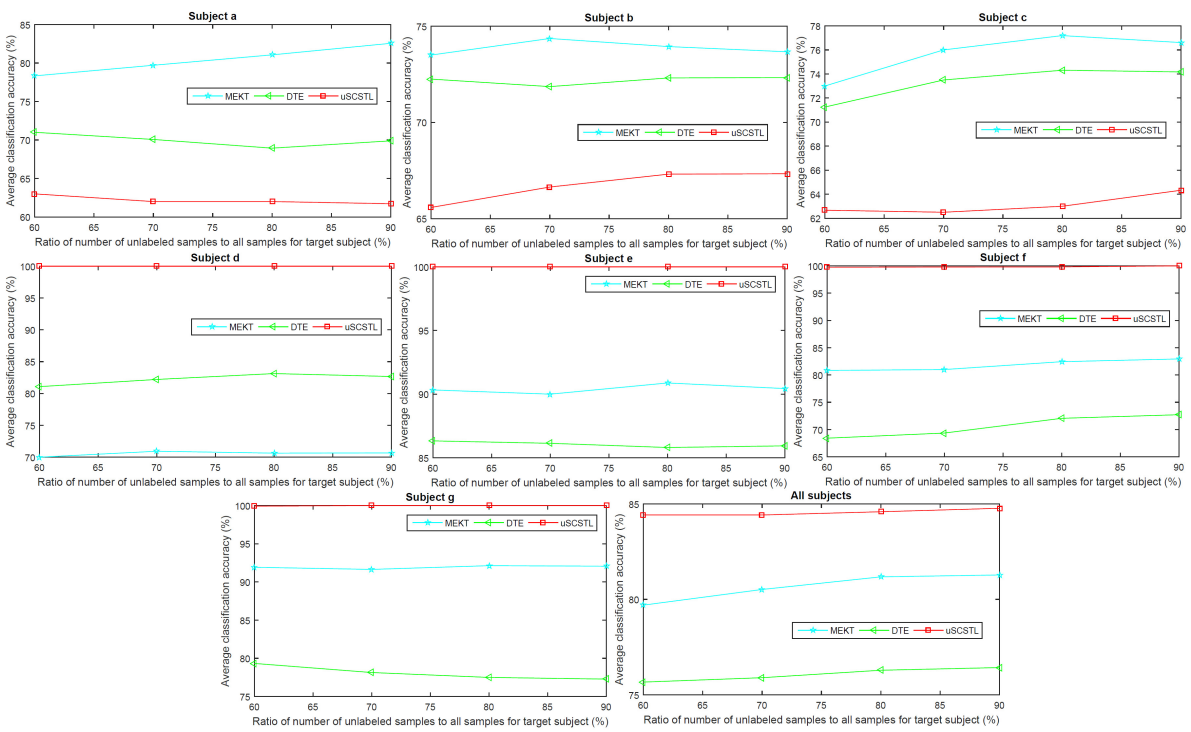


FIGURE 7 | Average classification accuracies of all subjects on dataset MI2, and their means, with varying R_u, for unsupervised algorithms.

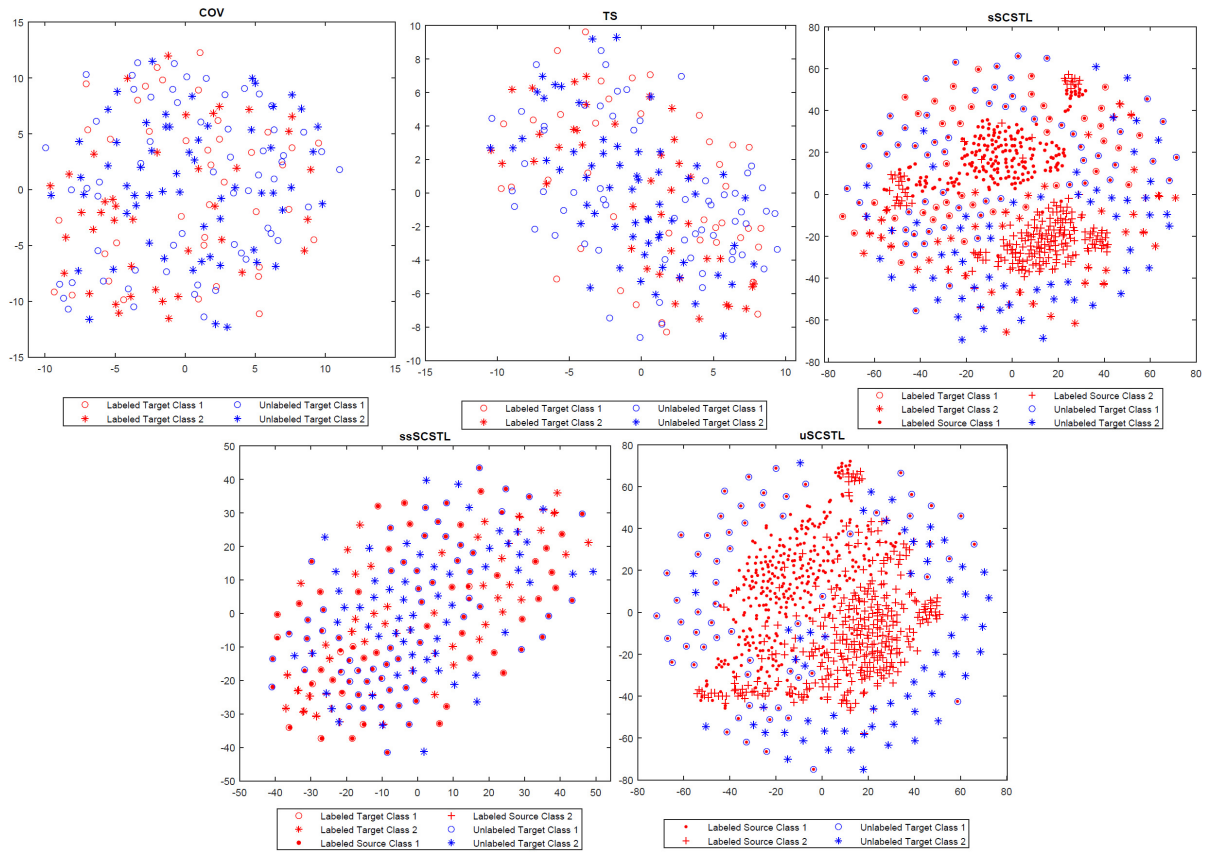


FIGURE 8 | The feature distributions of different algorithms displayed in the two-dimensional space.

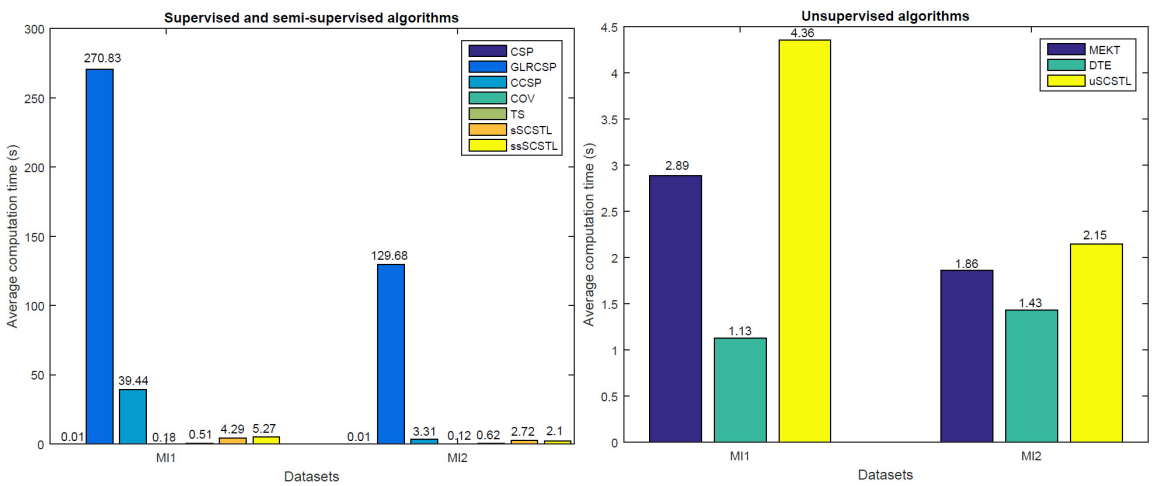


FIGURE 9 | Computation time comparison of different algorithms.

The baseline algorithms and our proposed algorithms are divided into three categories according to their feature spaces, including Euclidean space-based, Riemannian space-based, and Riemannian tangent space-based algorithms. For Euclidean space-based cross-subject TL algorithms, GLRCSP and CCSP spend much time in computing optimal regularization parameters by means of cross-validation to weight the labeled samples from different subjects. For Riemannian tangent space-based algorithms, the hyper-parameters of MEKT and DTE are set by experience. Our proposed algorithms are parameter

free. As shown in **Figure 8**, compared with Riemannian space-based algorithm COV, our proposed algorithms can make the samples from same class close to each other. In a word, our proposed algorithms can balance the classification performance and computation time well.

(3) In terms of different subjects

All subjects on MI1 or MI2 are separated into good and bad subjects according to their classification performances. To avoid negative transfer, most suitable source subjects are selected from good subjects for our proposed algorithms. As shown in **Figures 4**-**7**, our proposed algorithms exhibit obvious improvements over other algorithms for good target subjects. However, our proposed algorithms perform unsteadily for bad target subjects. The reason is that the classification performances of our proposed algorithms depend on the discriminability of the two-class target samples. For good target subjects, high discriminability of labeled or unlabeled target samples can help our proposed algorithms shorten the distances of samples with same class and select most appropriate source subjects. Nevertheless, for bad target subjects, low discriminability of labeled or unlabeled target samples influences the aggregation of the samples with same class and the source selection.

(4) In terms of labeled and unlabeled samples

To investigate the role of labeled and unlabeled samples, we propose sSCSTL, ssSCSTL, and uSCSTL. The labeled, total, or unlabeled samples from target subject are used to compute the Riemannian mean and source selection. As given in **Tables 1**, **5**, with the help of unlabeled samples from target subject, ssSCSTL performs slightly better than sSCSTL. Possible explanation is that adding unlabeled samples from target subject can not only make all samples from target subject closer, but also be good for better source selection. As shown in **Tables 2**, **6**, our proposed uSCSTL displays better classification performance than MEKT and DTE for good subjects. It suggests that unlabeled samples from good target subject which have inherent high discriminability can boost the performance improvement even if labeled samples from target subject are scarce.

(5) In terms of different classifiers

The baseline algorithms and our proposed algorithms use different supervised classifiers. Even for unsupervised TL algorithms, MEKT, DTE, and uSCSTL only use the unlabeled samples from target subject in the feature extraction phase. Except for TS, all Riemannian tangent space-based algorithms use sLDA as classifier. From the experimental results mentioned above, sLDA is much suitable for high-dimensional features. All Euclidean space-based algorithms and TS use LDA as classifier. As shown in **Tables 1**, **5**, CSP outperforms TS in most cases. Possible explanation is that LDA may not cope with high-dimensional features of TS well. Although the MDM classifier is much suitable for the classification of covariance matrices, the COV algorithm does not perform ideally due to without RA.

CONCLUSION

In this paper, we propose selective cross-subject transfer learning in supervised, semi-supervised, and unsupervised versions based on Riemannian tangent space. They perform RA using the available samples from target subject and the labeled samples from source subjects to preliminarily reduce the inter-subject differences. Then, all aligned covariance matrices from different subjects are converted into corresponding tangent space vectors for classification in Euclidean space. To realize positive transfer and relieve computational cost, the available tangent space vectors from target subjects are used to choose most suitable good source subjects based on SFFS. Experimental results show that our proposed algorithms are superior to several state-of-the-art algorithms, especially for good target subjects, when they have few or no labeled samples. Therefore, our proposed algorithms provide a new idea of solution for shortening calibration time in MI BCI.

However, our proposed ssSCSTL and uSCSTL algorithms are designed offline since the unlabeled samples from target subject are obtained a *prior*, instead of on-the-fly. Thus, we will further investigate how to adapt ssSCSTL and uSCSTl to the real-time BCI applications. Besides, ssSCSTL and uSCSTL only utilize the unlabeled samples from target subject in the feature extraction phase. Future work will be dedicated to extending ssSCSTL and uSCSTL in the classification phase to make full use of the unlabeled samples from target subject.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Dataset MI1: http://dx.doi.org/10.5524/100295, and Dataset MI2: http://www.bbci.de/competition/iv/desc_1.html.

AUTHOR CONTRIBUTIONS

YX wrote the manuscript, performed the research, and analyzed the experimental data. XH and QL designed, reviewed, and edited the manuscript. All authors read and approved the submitted manuscript.

FUNDING

This work was partially funded by the National Natural Science Foundation of China (Grant No. 62166020) and the Science and Technology Research Project of Jiangxi Provincial Department of Education (Grant No. GJJ200318).

ACKNOWLEDGMENTS

The authors would like to thank Dr. Ronghua Hu and the referees for their constructive recommendations, which help to improve the manuscript substantially.

REFERENCES

Azab, A., Ahmadi, H., Mihaylova, L., and Arvaneh, M. (2020). Dynamic time warping-based transfer learning for improving common spatial patterns in brain-computer interface. *J. Neural Eng.* 17:016061. doi: 10.1088/1741-2552/ab64a0

Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012). Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Trans. Biomed. Eng.* 59, 920–928. doi: 10.1109/TBME.2011.2172210

Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing* 112, 172–178. doi: 10.1016/j.neucom.2012.12.039

Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., et al. (2016). The Berlin brain-computer interface: progress beyond communication and control. *Front. Neurosci.* 10:530. doi: 10.3389/fnins.2016.00530

Blankertz, B., Dornhege, G., and Krauledat, M. (2007). The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *Neuroimage* 37, 539–550. doi: 10.1016/j.neuroimage.2007.01.051

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP componentsa tutorial. *NeuroImage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Cheng, M., Lu, Z., and Wang, H. (2017). Regularized common spatial patterns with subject-to-subject transfer of EEG signals. *Cogn. Neurodyn*. 11, 173–181. doi: 10.1007/s11571-016-9417-x

Cho, H., Ahn, M., Ahn, S., Kwon, M., and Jun, S. C. (2017). EEG datasets for motor imagery brain computer interface. *GigaScience* 6, 1–8. doi: 10.1093/gigascience/gix034

Fletcher, P. T., and Joshi, S. (2004). "Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors," in *Proceedings of the 2004 Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, Prague, 87–98. doi: 10.1007/978-3-540-27816-0_8

He, B., Baxter, B., Edelman, B. J., Cline, C. C., and Ye, W. W. (2015). Noninvasive brain-computer interfaces based on sensorimotor rhythms. *Proc. IEEE* 103, 907–925. doi: 10.1109/JPROC.2015.2407272

He, H., and Wu, D. (2020). Transfer learning for brain-computer interfaces: a Euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* 67, 399–410. doi: 10.1109/TBME.2019.2913914

Huang, C., Huang, X., Fang, Y., Xu, J., Qu, Y., Zhai, P., et al. (2020). Sample imbalance disease classification model based on association rule feature selection. *Pattern Recogn. Lett.* 133, 280–286. doi: 10.1016/j.patrec.2020.03.016

Huang, X., Yan, F., Xu, W., and Li, M. (2019). Multi-attention and incorporating background Information model for chest X-ray image report generation. *IEEE Access* 7, 154808–154817. doi: 10.1109/access.2019.2947134

Kang, H., Nam, Y., and Choi, S. (2009). Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Process. Lett.* 16, 683–686. doi: 10.1109/LSP.2009.2022557

Liang, Y., and Ma, Y. (2020). Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning. *Biomed. Signal Proces.* 62:102101. doi: 10.1016/j.bspc.2020.102101

Lotte, F., and Guan, C. (2011). Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* 58, 355–362. doi: 10.1109/TBME.2010.2082539

Lu, H., Plataniotis, K. N., and Venetsanopoulos, A. N. (2009). "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proceedings of the 2009 IEEE Engineering in Medicine and Biology Society* (Minneapolis, MN: IEEE), 6599–6602. doi: 10.1109/iembs.2009.5332554

Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* 26, 735–747. doi: 10.1137/s0895479803436937

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Peck, R., and Van Ness, J. (1982). The use of shrinkage estimators in linear discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 4, 530–537. doi: 10.1109/tpami.1982.4767298

Pudil, P., Ferri, F. J., and Kittler, J. (1994). "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proceedings of the 12th International Conference on Pattern Recognition* (Jerusalem, Israel: IEEE), 279–283. doi: 10.1109/ICPR.1994.576920

Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* 8, 441–446. doi: 10.1109/86.895946

Rodrigues, P. L. C., Jutten, C., and Congedo, M. (2019). Riemannian procrustes analysis: transfer learning for brain–computer interfaces. *IEEE Trans. Biomed.Eng.* 66, 2390–2401. doi: 10.1109/TBME.2018.2889705

Sample, M., Aunos, M., Blain-Moraes, S., Bublitz, C., Chandler, J. A., Falk, T. H., et al. (2019). Brain–computer interfaces and personhood: interdisciplinary deliberations on neural technology. *J. Neural Eng.* 16:063001. doi: 10.1088/1741-2552/ab39cd

Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6:55. doi: 10.3389/fnins.2012.00055

Tuzel, O., Porikli, F., and Meer, P. (2008). Pedestrian detection *via* classification on Riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.* 30, 1713–1727. doi: 10.1109/TPAMI.2008.75

Van Der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Xu, Y., Wei, Q., Zhang, H., Hu, R., Liu, J., Hua, J., et al. (2019). Transfer learning based on regularized common spatial patterns using cosine similarities of spatial filters for motor-imagery BCI. *J. Circuits Syst. Comp.* 28, 1950123. doi: 10.1142/S0218126619501238

Yin, E., Zeyl, T., Saab, R., Chau, T., Hu, D., and Zhou, Z. (2015). A hybrid brain–computer interface based on the fusion of P300 and SSVEP scores. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 693–701. doi: 10.1109/TNSRE.2015.2403270

Zanini, P., Congedo, M., Jutten, C., Said, S., and Berthoumieu, Y. (2018). Transfer learning: a Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 65, 1107–1116. doi: 10.1109/TBME.2017.2742541

Zhang, W., and Wu, D. (2020). Manifold embedded knowledge transfer for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1117–1127. doi: 10.1109/TNSRE.2020.2985996

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

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

Copyright © 2021 Xu, Huang 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 in Neuroscience

ORIGINAL RESEARCH
published: 16 November 2021
doi: 10.3389/fnins.2021.756536

Voxel-Wise Adversarial FiboNet for 3D Cerebrovascular Segmentation on Magnetic Resonance Angiography Images

Bin Guo, Fugen Zhou, Bo Liu and Xiangzhi Bai**

Image Processing Center, School of Astronautics, Beihang University, Beijing, China

Cerebrovascular segmentation is important in various clinical applications, such as surgical planning and computer-aided diagnosis. In order to achieve high segmentation performance, three challenging problems should be taken into consideration: (1) large variations in vascular anatomies and voxel intensities; (2) severe class imbalance between foreground and background voxels; (3) image noise with different magnitudes. Limited accuracy was achieved without considering these challenges in deep learning-based methods for cerebrovascular segmentation. To overcome the limitations, we propose an end-to-end adversarial model called FiboNet-VANGAN. Specifically, our contributions can be summarized as follows: (1) to relieve the first problem mentioned above, a discriminator is proposed to regularize for voxel-wise distribution consistency between the segmentation results and the ground truth; (2) to mitigate the problem of class imbalance, we propose to use the addition of cross-entropy and Dice coefficient as the loss function of the generator. Focal loss is utilized as the loss function of the discriminator; (3) a new feature connection is proposed, based on which a generator called FiboNet is built. By incorporating Dice coefficient in the training of FiboNet, noise robustness can be improved by a large margin. We evaluate our method on a healthy magnetic resonance angiography (MRA) dataset to validate its effectiveness. A brain atrophy MRA dataset is also collected to test the performance of each method on abnormal cases. Results show that the three problems in cerebrovascular segmentation mentioned above can be alleviated and high segmentation accuracy can be achieved on both datasets using our method.

Keywords: index terms-class imbalance, image noise, CNN, GAN, medical image segmentation

OPEN ACCESS

Edited by:
Kaijian Xia,
Changshu No.1 People's Hospital, China

Reviewed by:
Yu Liu,
Hefei University of Technology, China
Pramod Kumar Pisharady,
University of Minnesota, United States
Wei Zhang,
University of California, San Francisco, United States
Deepti R. Bathula,
Indian Institute of Technology Ropar, India

****Correspondence:***
Bo Liu
bo.liu@buaa.edu.cn
Xiangzhi Bai
jackybxz@buaa.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *10 August 2021*
Accepted: *13 October 2021*
Published: *16 November 2021*

Citation:
Guo B, Zhou F, Liu B and Bai X (2021) Voxel-Wise Adversarial FiboNet for 3D Cerebrovascular Segmentation on Magnetic Resonance Angiography Images. Front. Neurosci. 15:756536. doi: 10.3389/fnins.2021.756536

1. INTRODUCTION

Cerebrovascular diseases, such as strokes and aneurysms, are among the most important public health problem around the world. Although prevalence of lethal vascular diseases such as aneurysm is relatively low, estimated to be between 1 and 5% (Brisman et al., 2006), these diseases usually have a high fatality rate (McKinney et al., 2008). In order to find suitable predictors of risk of vascular diseases, computational modeling is increasingly used. Most notably, shape characterization and analysis of hemodynamic features of vessels are becoming increasingly important in prediction of aneurysm and stenosis (Raghavan et al., 2005; Millán et al., 2007). These results are strongly

determined by the modeled geometry of vessels. Therefore, accurate vascular segmentation is of vital importance.

Traditional machine learning-based methods (Hassouna et al., 2006; Soares et al., 2006; Oliveira et al., 2011; Mapayi et al., 2015; Goceri et al., 2017) include unsupervised learning methods and supervised learning methods. In unsupervised learning methods, feature models for target should be proposed to distinguish vessels from background. While in supervised learning methods, pixel-wise classification in training and testing is very time consuming. Also, feature engineering is unavoidable, which is tedious and requires expertise of target domain. Recent advances in deep learning have enabled training of complex methods including deep convolutional neural networks (CNNs). Various CNNs have been proposed (Lee et al., 2015; Xie and Tu, 2015; Dou et al., 2016; Milletari et al., 2016; Christ et al., 2017; Shelhamer et al., 2017; Yu et al., 2017) in medical image segmentation. These methods require no hand-tuned image features and can be plug and play on different dataset. However, different characteristics usually exist in different dataset. If these dataset-related problems are considered in the designing and training of CNNs, higher segmentation accuracy can be realized.

Actually, in medical image segmentation, three main challenges exist. First, large variations of anatomies (Nain et al., 2004; Zheng et al., 2012, 2014; Lugauer et al., 2014; Zheng, 2016) and voxel intensities exist (Iqbal, 2013). For cerebrovascular segmentation using MRA images, large variation of vascular radius exists among different vascular branches (Rätsep et al., 2016). This variation usually corresponds to the large variation of blood flow rate among vessels (Zarrinkoob et al., 2015). Due to the imaging principle of MRA where higher blood flow rate will result in higher voxel intensity, large variations of voxel intensities are also obvious among vessels. Second, severe class imbalance makes the learning-based methods converge to biased local minima (Milletari et al., 2016; Buda et al., 2018), leading to compromised performance. In our dataset, less than 0.3% voxels belong to vascular regions. Third, image noise is an unavoidable phenomenon (Li et al., 2017), which deteriorates the quality of images and further suppresses the segmentation performance of various methods.

In this paper, we take the above problems of cerebrovascular segmentation into consideration for network designing and training. A model called voxel-wise adversarial FiboNet, termed as FiboNet-VANGAN, is tailored to relieve these problems. Specifically, our contributions are summarized as follows:

- Adversarial training is incorporated and a voxel-wise adversarial network is proposed to relieve the problem of large variations in vascular anatomies and voxel intensities.
- In order to relieve the problem of class imbalance, we propose to use the addition of cross-entropy and Dice coefficient (DC) as the loss function of the generator. Moreover, focal loss is employed as the loss function (FL) of discriminator to relieve the problem of class imbalance.
- We propose a new feature aggregation-based generator called FiboNet. By applying DC as one of the loss functions of FiboNet, the noise robustness of FiboNet is improved.

Experimental results validate the effectiveness of our contributions and segmentation results with high accuracy are achieved by our proposed model on the Healthy Dataset and the Brain Atrophy Dataset.

2. RELATED WORKS

Popular networks, such as U-Net (Ronneberger et al., 2015) and V-Net (Milletari et al., 2016) can be adapted to carry out cerebrovascular segmentation. However, these networks were initially proposed to deal with organ segmentation. In organ segmentation, target organs usually tightly distribute in an image volume, which is quite different from the cerebrovascular segmentation. The backbones of these two networks incorporate many max-pooling layers (four max-pooling layers are adopted in both networks) on the encoding path. Hence, networks can encode more multi-scale informative features. On the decoding path, corresponding de-convolutional layers are included to decode these features back to the scale of input image. Both of these networks use image of the full size (or cropped sub-image) as the inputs to train the networks. However, in cerebrovascular segmentation, the size of each MRA volume is much larger than the dataset of U-Net and V-Net, making it improper to apply the image of full size as the input to train networks due to limited graphic memory. Also, unlike the distribution of organs, vessels are thin and distribute sparsely and extensively within the brain. The smallest sub-image we can extract is the bounding box of the brain, which is still quite large. Therefore, to apply these two networks on cerebrovascular segmentation, patch-wise training strategy is adopted. Actually, in cerebrovascular segmentation, patch-wise training strategy may be regarded as an implicit method of data augmentation. From a global point of view, the whole cerebral vasculatures are quite different with each other intersubjectively. It is difficult for networks to learn the distribution of the whole cerebral vasculature using limited number of data. But when it comes to a local point of view, partial cerebral vasculatures are likely to share similar interhemispheric and intersubjective features, which makes it very suitable to apply patch-wise training strategy.

Aside from these two networks, several CNN-based methods have been proposed for vascular segmentation (Merkow et al., 2016; Chen et al., 2017a; Yu et al., 2017; Tetteh et al., 2018). In Merkow et al. (2016), they derived their 3D vascular network from a multi-scale 2D network for detection of boundaries. The technique called deep supervision (Lee et al., 2015) is combined and is placed at each scale to improve the performance of their network. However, they proposed no methods to deal with any of the challenges mentioned above. Hence, only limited results can be achieved. In Yu et al. (2017) adopted dense connection (Huang et al., 2017) as the backbone of their network. This type of feature aggregation is deemed to promote the training of networks. Hence, compared with networks without dense connection, their network can achieve better segmentation results. In Tetteh et al. (2018) discussed the problem of class imbalance in this application. A class balancing loss function was employed to train their network. But they did not deal with the other two problems.

Chen et al. (2017a) proposed to use convolutional autoencoder to reduce the influence of noise. But they mentioned no methods to mitigate the problem of class imbalance. None of the above networks took the large variations of vascular anatomies and voxel intensities into consideration.

Actually, the problem of anatomical variation also occurs in other vascular systems (Nain et al., 2004; Zheng et al., 2012, 2014; Lugauer et al., 2014; Zheng, 2016). Data-driven approaches (Lesage et al., 2009; Schaap et al., 2009) are capable of addressing the problem of anatomical variation due to their intrinsic bottom-up paradigm. But this type of algorithm may terminate early since no or little high-level prior information is used (Zheng et al., 2013). In Zheng et al. (2014), a part based model is proposed. The anatomical variations of the pulmonary vein (PV), e.g., the left common PVs vs. separate PVs, can be addressed using this model. Similar model-driven methods are also applied to address the anatomical variations in segmenting coronary arteries (Nain et al., 2004; Zheng et al., 2012; Lugauer et al., 2014). Zheng (2016) also proposed a hybrid method in segmenting coronary arteries. Model-based approach exploiting shape priors is utilized to extract the major coronary arteries, followed by a data-driven approach to address the anatomical variations in side branches.

3. METHODOLOGY

In section A, we present our novel CNN-based generator (G) called FiboNet. The derivation of Fibonacci connection will be described. In section B, we propose an adversarial network as the discriminator (D). D is adopted as the regularizer for voxel-wise distribution consistency between the predictions from FiboNet and the ground truth. Training procedure and loss functions will be detailed in section C.

3.1. FiboNet as Generator

The proposed FiboNet is illustrated in **Figure 1**. In CNNs, shallow layers are responsible for extracting low level features, such as edges and curves of different orientations and sizes, while deep layers are responsible for generating semantic information (Krizhevsky et al., 2012). If low level features are similar with each other, only limited kinds of semantic information can be generated by deep layers. Therefore, if a convolutional block can learn to extract diversified feature maps from a certain input image, very different high-level features can be represented. In the problem of cerebrovascular segmentation, different high-level features can help relieve the problem of large variation of vascular anatomies and voxel intensities.

3.1.1. Fibonacci Connection and Fibonacci Block

Fibonacci connection is derived from the dense connection (Huang et al., 2017). Dense connection can be formulated as:

$$x_l = H_l([x_{l-1}, x_{l-2}, ..., x_0]) \tag{1}$$

where x_i represents feature maps generated by the ith layer. $H_l(\cdot)$ represents a composite function from l^{th} layer, including batch normalization (Ioffe and Szegedy, 2015), rectified linear unit (Nair and Hinton, 2010), and convolution. By this formulation, the constraints from the current layer will directly influence each of its previous layers. Similar feature maps will be generated by each convolution layer within the same dense block (Chen et al., 2017b) due to its strong regularizing effects (Dolz et al., 2018). This may result in similar feature maps shared across different layers within the same dense block. In cerebrovascular segmentation, large variations of vascular anatomies and voxel intensities exist. Hence, we are desiring for diversified features from a certain input image. We achieve this goal by weakening the strong regularizing effect of the dense connection. We cut off some of the feature connections so that constraints from the current layer cannot directly influence each of its previous layers. Specifically, in our case, we concatenate the feature maps from previous two layers only, which is formulated as:

$$x_l = H_l([x_{l-1}, x_{l-2}]) \tag{2}$$

In this type of feature aggregation, only the previous two layers are directly affected by current layer. This type of feature aggregation coincides with the developing manner of Fibonacci numbers. Therefore, we term this type of feature aggregation as Fibonacci connection. In Fibonacci connection, layers before $l-1, l-2$ are indirectly influenced. The regularizing effect of the dense block is weakened. Since diversified feature maps are extracted by this type of feature connection, it can help detect more kinds of vascular voxels.

3.1.2. FiboNet

Using Fibonacci connection, we propose the FiboNet, as illustrated in **Figure 1**. In order to detect the thin boundaries of vessels, small kernel size is used in the whole network. In shallow layers of the FiboNet, we are expecting for weak regularizing effect between different layers in order to generate diversified feature maps and mitigate the under-detection of candidate vascular voxels.

Hence, we use the Fibonacci block in shallow layers. While dense block is placed in deep layers to extract similar high-level features belonging to vascular voxels and to avoid as much over-detection of background voxels as possible. Notably, we included a branch at the output of FiboNet as deep supervision (Lee et al., 2015), allowing gradients to inject directly into the preceding layers to better train the network (Lee et al., 2015).

3.2. Voxel-Wise Adversarial Network as Discriminator

In cerebrovascular segmentation, large variations exist in vascular anatomies and voxel intensities within even the same subject. In MRA images, signals with high intensity spread in the vessels with large radius. Therefore, signals with low intensity are comparatively harder to be segmented because signals with lower intensity are contained within smaller vessels, which consist of fewer voxels. Even though patch-wise training strategy is adopted, the influence of these voxels is easy to get overwhelmed by background voxels. Since a thorough knowledge of anatomy is required to understand and analyze the cerebrovascular system (Nowinski et al., 2009) with respect to different clinical purposes,

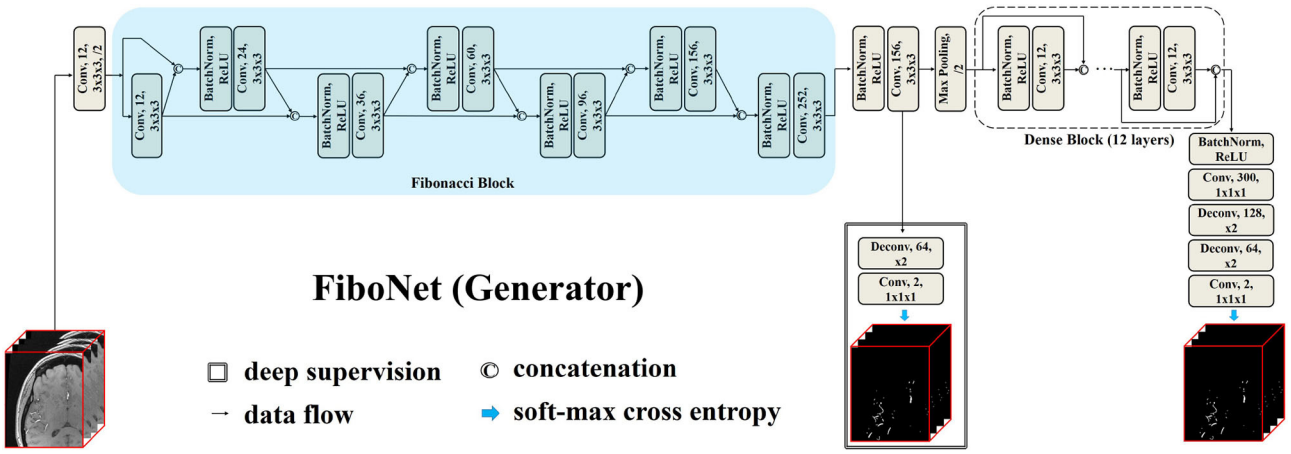


FIGURE 1 | Architecture of our proposed FiboNet as the generator. Fibonacci connection of feature maps are utilized to build the Fibonacci block. In Fibonacci block, feature maps from previous two layers are concatenated as the input of the current layer. The technique called deep supervision, which is commonly used in convolutional neural network (CNN)-based medical image segmentation, is also incorporated.

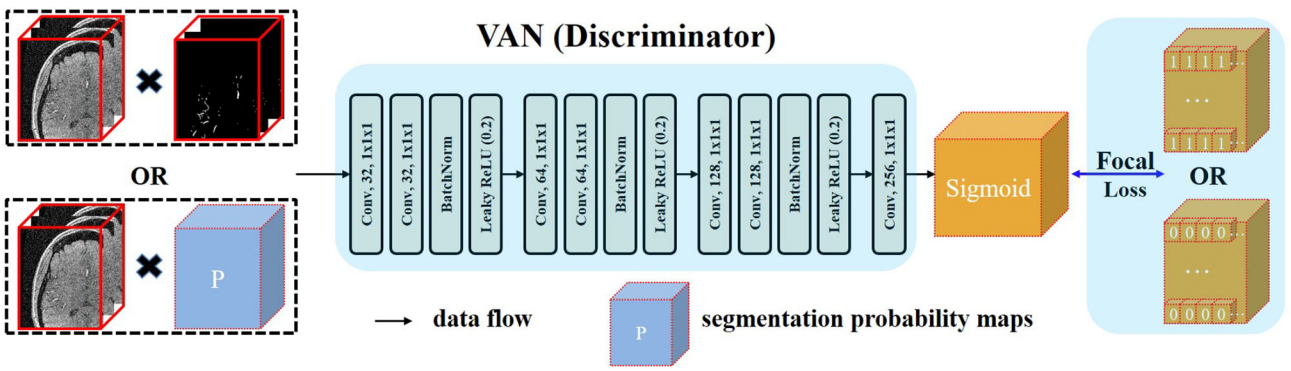


FIGURE 2 | Architecture of the proposed voxel-wise adversarial network (VAN). VAN is trained to distinguish the source of its input. In the context of generative adversarial network (GAN), the procedure of adversarial training is to minimize the discrepancy between two distributions. Since we want voxel-wise similar distribution between the distribution generated by FiboNet and the real data distribution, we input these two types to VAN. Kernel size of 1 is used in VAN. FL is utilized as the loss function to relieve the problem of class imbalance.

such as automatic diagnosis and surgical planning, it is important to improve the segmentation performance of networks within these regions.

Recent development of deep learning has validated the effectiveness of GAN (Goodfellow et al., 2014; Luc et al., 2016) in regularizing higher order consistency (Luc et al., 2016) including the category consistency between generated images and images from the training dataset (Goodfellow et al., 2014) and the texture consistency between generated images and real images (Isola et al., 2017). GAN can be regarded as a training framework. In this framework, the generator (G) and the discriminator (D) are jointly trained. This training procedure will make the distribution generated by G close to the real data distribution. In Goodfellow et al. (2014), whole image is input to the D and D outputs one digit to represent the source of its input. This type of adversarial training was usually applied in regularizing category consistency. Isola et al. (2017) proposed another adversarial strategy by constraining the receptive field of each output voxel of D to a patch of $D's$ input. By this operation, they can impose texture consistency between generated images and real images. In our case, we embrace the similar idea and propose a voxel-wise adversarial network, termed as voxel-wise adversarial network (VAN), to incorporate voxel-wise distribution consistency by constraining the receptive field of each output voxel of D to a voxel of $D's$ input. Using this type of adversarial training, the segmentation performance of networks in small vessel regions can be consistently improved.

Details of the architecture of VAN are illustrated in **Figure 2**. Our VAN accepts two types of inputs: one is the multiplication of noisy raw patches and the segmentation probability maps; the other is the multiplication of noisy raw patches and the annotated ground truth patches. Noisy raw patches are used as part of $D's$ input in case of the degenerate distributions generated by G (Sønderby et al., 2016). Multiplication operation

is used so that VAN can take the relationships of raw patches and the predictions/ground truth into consideration from the beginning. Thus, all the model parameters are jointly trained using information from both the segmentation and raw patches.

3.3. Training and Losses

The training pipeline of the proposed FiboNet-VANGAN is illustrated in **Figure 3**. The following two stages are repeated to iteratively train the generator and discriminator.

3.3.1. First Stage of Training

In this stage, G is trained to generate segmentation probability maps close to the ground truth and also is encouraged to fool D about the source of its input. The loss function L_G is formulated as:

$$L_G = \sum_{i=1}^{N} l_{CE}(P_i, Y_i) - l_{DC}(P_i, Y_i) - \sum_{i=1}^{N} l_{FL}(S_i = D(X_i \cdot P_i), T_i = 0) \tag{3}$$

where N represents the number of input patches, P_i represents the generated probability map of $G's$ softmax layer, and Y_i represents the corresponding ground truth. S_i represents the predicted score maps of D, and T_i represents target score maps of D. CE represents cross-entropy, and DC represents Dice

Coefficient FL represents focal loss and these three losses can be formulated as:

$$l_{CE}(P_i, Y_i) = -\sum_{c=0}^{1}\sum_{v}^{V} y_{c,v} log(p_{c,v}) \tag{4}$$

$$l_{DC}(P_i, Y_i) = \sum_{c=0}^{1} \frac{2\sum_{v}^{V} p_{c,v} y_{c,v}}{\sum_{v}^{V} p_{c,v}^2 + \sum_{v}^{V} y_{c,v}^2} \tag{5}$$

$$l_{FL}(S_i, T_i) = (1 - S_i)^{\gamma} \cdot l_{CE}(S_i, T_i) \tag{6}$$

where V represents the number of voxels in current patch P_i, $y_{c,v}$ represents target probability of vth voxel belonging to c^{th} class, and $p_{c,v}$ represents generated probability of vth voxel belonging to c^{th} class. In practice, one-hot encoding is adopted. Hence, $y_{c,v} = 1$ is used. γ is a decaying factor and $\gamma = 4$ is used in our implementation.

CE enforces voxel-wise similarity between predicted distribution of G and real data distribution. FiboNet trained with only CE tends to get trapped in biased local minima (Milletari et al., 2016) with the existence of severe class imbalance, because voxel-wise penalization is very sensitive to class imbalance. To handle this, we propose to use the addition of CE and DC as the loss function. As DC will incur penalization on overlapping areas, which can be concluded from Equation (5), the problem of class imbalance could be alleviated. When predictions completely overlap with ground truth, DC will output 1. When there are no intersections between predictions and ground truths, DC will output 0. In practice, we minimize $(1 - DC)$.

FL is a weighted version of traditional binary cross entropy. In our case, voxel-wise adversarial training is employed to model steep regions in which D is trained to distinguish the source of its input voxel-wisely. Therefore, the problem of class imbalance will also have an influence on the training of D. Thus, we should take this problem into consideration in the adversarial training procedure of both G and D. During the training of D, it is desired that major components of loss values comes from those falsely distinguished voxels. Due to the severe class imbalance, D will learn quickly to distinguish the source of background voxels. Under this circumstance, loss values from foreground voxels are desired to be the major component of the total loss. But the number of these voxels are too small. Their total loss is overwhelmed by the total loss from the background voxels though these background are truely distinguished. Therefore, we need to lower the contributions of those correctly distinguished voxels. In Lin et al. (2017), the FL is proposed to improve the detection accuracy of single-stage detectors. Inspired by their work, we apply FL as the loss function of our VAN to relieve the problem caused by class imbalance. With this loss function, when D can output scores close to target scores of each background voxel, the weighted factor will decrease their contributions to the total loss. Thus, loss values from falsely distinguished voxels will be the major component.

3.3.2. Second Stage of Training

In this training stage, D is trained to distinguish the sources of its input voxel-wisely. Two types of inputs are accepted. One is the multiplication of the segmentation probability maps and noisy raw patch. The other is the multiplication of the ground truth and raw patch added with noise. When the first one is input into D, it is desired to be recognized as 0 voxel-wisely by D. When the second one is input into D, it is desired to be recognized as 1 voxel-wisely by D. At this training stage, we use the FL (Lin et al., 2017) as the loss function, which is formulated as:

$$L_D = \sum_{i=1}^{N} l_{FL}(S_i = D(\hat{X}_i \cdot P_i), T_i = 0) + \sum_{i=1}^{N} l_{FL}(S_i = D(\hat{X}_i \cdot Y_i), T_i = 1) \tag{7}$$

where $\hat{X}_i$ represents the noisy raw patch. Noisy patch is used as part of $D's$ input in case of degenerate distribution generated by G (Sønderby et al., 2016). In the beginning of training G, a good prediction result cannot be achieved. Obvious differences exist between P_i and Y_i. Therefore, the two predictions, $D(\hat{x}_i \cdot P_i)$ and $D(\hat{x}_i \cdot Y_i)$, of D are also quite different. Under this situation, the weighting factor of FL does not play a very important role. But when a large number of background voxels can be truely predicted by G, P_i and Y_i will be very similar. In this case, loss values calculated by D will be biased since we are applying adversarial training voxel-wisely. When G is trained with the loss value from biased D, the accuracy of G may drop. But the weighting factor in FL can make a balance between the foreground and the background voxels with the occurrence of class imbalance. It can lower the contributions of background

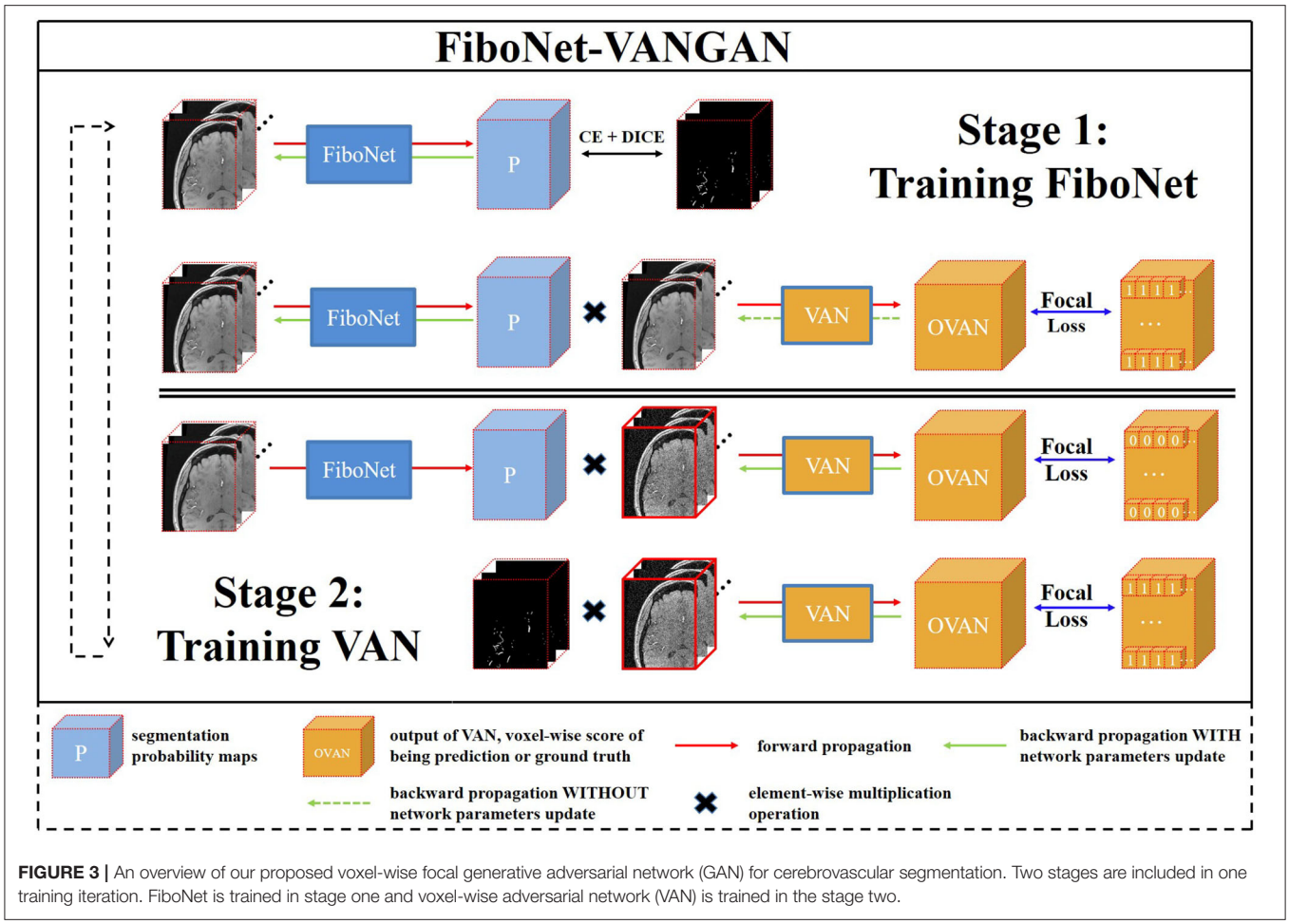


FIGURE 3 | An overview of our proposed voxel-wise focal generative adversarial network (GAN) for cerebrovascular segmentation. Two stages are included in one training iteration. FiboNet is trained in stage one and voxel-wise adversarial network (VAN) is trained in the stage two.

voxels. Thus, the problem of biased convergence can be relieved by a large extent.

4. EXPERIMENTS AND RESULTS

4.1. Datasets and Preprocessing

Two datasets, acquired on a 1.5 T GE BRIVO MR355 using gradient echo sequence with repetition time of 26 ms and echo time of 6.8 ms, are collected and used for method evaluation in this work, the healthy dataset and the brain atrophy dataset. Procedures of data collections were reviewed by Datian County Hospital Ethics Committee. Participants provided informed consent to participate in the study. Two clinicians were responsible for the annotation of ground truths. A consensus between them was used as the final reference standard.

4.1.1. Healthy Dataset

This dataset contains 12 healthy time-of-flight MRA (TOF-MRA) volumes. Each volume is reconstructed into a matrix of size $1,024 \times 1,024 \times 92$ with voxel size of $0.264 \times 0.264 \times 0.8mm$. No contrast agent is used during the scanning. No removal of bias field is carried out. The age and gender distribution of this dataset are listed in **Table 1**.

TABLE 1 | The age and gender distribution of the healthy dataset.

	Healthy dataset			
Gender \ Age	(20,30]	(30, 40]	(40, 50]	Total
Male	3	1	1	5
Female	1	3	3	7
Total	4	4	4	12

4.1.2. Brain Atrophy Dataset

This dataset was collected to test the performance of each network on abnormal dataset, contains 9 TOF-MRA volumes from 9 subjects diagnosed with brain atrophy. Each volume within this dataset is reconstructed into a matrix of $512 \times 512 \times 128$ with voxel size of $0.43 \times 0.43 \times 0.7mm$. The age and gender information of this group is listed in **Table 2**. Atrophy of any tissue within the brain means a decrement in the size of its cells. Atrophy can be generalized, which means that all of the brain has shrunk; or it can be focal, affecting only a limited area of the brain and resulting in a decrease of the functions that area of the brain controls. In either case, the distribution of the cerebral vasculature becomes quite different from the

TABLE 2 | The age and gender distribution of the brain atrophy dataset.

Brain atrophy dataset			
Gender \ Age	(70,80]	(80, 90]	Total
Male	4	0	4
Female	2	3	5
Total	6	3	9

healthy subjects even within the same range of ages. Therefore, this dataset makes a good testbed for the performance of each network on abnormal dataset.

4.1.3. Data Normalization

Before training and testing, we carry out the procedure of data normalization using the following equation:

$$X = \frac{X - mean(X)}{std(X)} \tag{8}$$

where X represents each MRA volume.

4.1.4. Data Splitting and Patch-Wise Training/Testing

The MRA volumes are too large to fit in our graphic memory during training and testing. Also, cerebral vessels usually share similar interhemispheric and intersubjective features. Therefore, we turn to patch-wise training strategy, which has been frequently used in the context of medical image segmentation (Merkow et al., 2016; Yu et al., 2017; Gibson et al., 2018; Han et al., 2018; Kushibar et al., 2018; Tetteh et al., 2018; Wang et al., 2018). Each volume from the healthy dataset is split into overlapping patches with size of $64 \times 64 \times 64$. An overlapping size of $4 \times 4 \times 36$ is adopted, thus corresponding to 17, 17, and 2 patches in each direction and 578 patches in total. For each volume from the brain atrophy dataset, the same patch size is used while overlapping sizes are changed to $4 \times 4 \times 32$ to avoid zero-padding in the axial direction, thus corresponding to 9, 9, and 3 patches in each direction and 243 patches for each MRA volume.

In the training stage, each MRA volume is split into patches to train the networks. In the testing stage, each volume is first split into corresponding patches. Then the patches are input to the networks. After the processing of networks, predicted patches are stitched into the dimension of the original testing volume. Major voting strategy is utilized to determine the categories of the voxels from overlapping regions.

All the networks are trained on the same 6 volumes, which are randomly selected from the healthy dataset. The rest 6 volumes of the healthy dataset are used as the testing dataset to test the segmentation performance of each network. The whole brain atrophy.

Dataset is utilized to evaluate the performance of each method on abnormal dataset. Thus, each network is trained on the same 6 volumes and tested on the same 15 volumes.

4.2. Experiment Setup

We compare our results with popular networks in 3D medical image segmentation including 3D-UNet, a 3D version of U-Net (Ronneberger et al., 2015), V-Net (Milletari et al., 2016), I2I-3D (Merkow et al., 2016), UCeption (Sanchesa et al., 2019), DenseVoxNet (Yu et al., 2017), and DeepVesselNet (Tetteh et al., 2018). We implement the derived 3D version of U-Net, and UCeption in tensorflow. As for other networks, original open source codes are used in the experiments.

In order to validate our proposal of using the addition of cross-entropy and DC, we also train both the DenseVoxNet and the FiboNet using addition of cross-entropy and DC as loss function (DenseVoxNet-CE-DC and FiboNet-CE-DC in **Table 3**).

In order to compare our Fibonacci connection with traditional linear convolutional (LinearCN) block where each layer only connects to its previous one, we also construct a LinearCN block by connecting each layer in Fibonacci block to its very preceding layer only.

VoxelGAN (the proposed VAN), PatchGAN, and VolumeGAN are three types of discriminators (Isola et al., 2017). In order to get a better understanding of the different influences from different adversarial training, we also carry out the experiments of different adversarial training by changing the proposed VAN to corresponding adversarial networks. Essentially, they are all fully convolutional networks but differ in the receptive field of the output voxel. When the receptive field corresponds to a voxel, it is termed as VoxelGAN. When the receptive field corresponds to a patch, it is termed as PatchGAN. When the receptive field corresponds to a volume, it is termed as VolumeGAN. In designing the architectures of PatchGAN and VolumeGAN, all convolutions are of kernel size 3 (or larger) with stride 2 (or larger) to expand the receptive field of voxels from follow-up layers. But in designing the architecture of VoxelGAN, kernel size 1 and stride 1 are adopted to restrict the receptive field. In our experiments, architectures of PatchGAN and VolumeGAN similar to Isola et al. (2017) are adopted:

- PatchGAN: $C64 - C128 - C256$
- VolumeGAN: $C64 - C128 - C256 - C512 - C512$

where Ck denotes a Convolution-BatchNorm-ReLU with k filters. All ReLUs are leaky with slope 0.2. The receptive field of a output voxel in PatchGAN is $15 \times 15 \times 15$ and VolumeGAN $63 \times 63 \times 63$.

All experiments were performed on workstation, which runs Ubuntu 16.04 operation system and is equipped with an Intel Core i7-5960 CPU (3.50 GHz), 32GB RAM, and a NVIDIA GeForce 1080 Ti video card with 11GB graphics memory. The SGD solver with learning rate 0.001, which exponentially decays by 10% every 5 epochs, and momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ are used for training all the adversarial networks. Each network is trained for 40 epochs to ensure convergence.

To quantitatively evaluate the performance of each method, three metrics are adopted to evaluate the segmentation accuracy:

- Dice coefficient (DC) Dice (1945):
  $DC = 2|P \cap G|/(|P| + |G|)$

TABLE 3 | Quantitative comparisons of different methods on different datasets.

Methods	Healthy dataset			Brain atrophy dataset		
	DC	***SASD*(*mm*)**	***SHD*95(*mm*)**	***DC***	***SASD*(*mm*)**	***SHD*95(*mm*)**
3D-UNet Ronneberger et al., 2015	0.7133 ± 0.0349	1.0447 ± 0.2428	7.4281 ± 1.3958	0.6478 ± 0.0264	1.5247 ± 0.4347	6.8728 ± 1.7996
V-Net Milletari et al., 2016	0.7255 ± 0.0320	0.7974 ± 0.1716	4.4059 ± 1.5024	0.5842 ± 0.0307	1.1460 ± 0.3138	6.8811 ± 2.1470
I2I-3D Merkow et al., 2016	0.7298 ± 0.0269	0.9267 ± 0.2282	6.3634 ± 1.8695	0.5979 ± 0.0321	1.3444 ± 0.5033	6.6167 ± 2.8714
DeepVesselNet Tetteh et al., 2018	0.7391 ± 0.0261	1.0805 ± 0.3400	6.6839 ± 1.1783	0.6188 ± 0.0282	1.6308 ± 0.5314	8.7423 ± 2.7256
UCeption Sanchesa et al., 2019	0.7680 ± 0.0250	0.7239 ± 0.1472	4.8261 ± 1.2030	0.6036 ± 0.0198	1.9772 ± 0.3509	12.2734 ± 1.3957
DenseVoxNet-CE Yu et al., 2017	0.7611 ± 0.0227	0.6523 ± 0.1296	4.1740 ± 1.4962	0.6524 ± 0.0086	1.4203 ± 0.2831	9.7092 ± 1.9996
DenseVoxNet-CE-DC	0.7840 ± 0.0266	0.8249 ± 0.2071	5.5383 ± 1.3207	0.7120 ± 0.0224	1.3878 ± 0.3092	9.5628 ± 1.9789
FiboNet-CE	0.7985 ± 0.0216	0.4699 ± 0.1110	2.7348 ± 1.2129	0.6985 ± 0.0110	1.0479 ± 0.2122	6.5655 ± 2.4639
FiboNet-CE-DC	0.8093 ± 0.0296	0.4848 ± 0.1447	2.9483 ± 1.2064	0.7532 ± 0.0190	1.3338 ± 0.2442	9.0029 ± 1.7674
LinearCN-CE-DC	0.8016 ± 0.0264	0.5032 ± 0.1098	3.0064 ± 1.0408	0.7292 ± 0.0194	1.1558 ± 0.2092	7.4721 ± 2.1517
FiboNet-PatchGAN	0.8188 ± 0.0269	0.4247 ± 0.1073	2.6115 ± 0.9309	0.7519 ± 0.0193	1.1599 ± 0.2472	7.9665 ± 2.2705
FiboNet-VolumeGAN	0.8153 ± 0.0278	0.4486 ± 0.1272	2.6436 ± 1.0850	0.7515 ± 0.0227	1.2192 ± 0.2588	8.2716 ± 2.1943
FiboNet-VANGAN	**0.8197 ± 0.0300**	**0.4019 ± 0.1111**	**2.4391 ± 0.9615**	**0.7571 ± 0.0173**	**1.0023 ± 0.2508**	**6.3991 ± 2.6469**

For the healthy dataset, results based on the rest 6 MRA volumes for testing are summarized. Bold values indicate best segmentation performance.

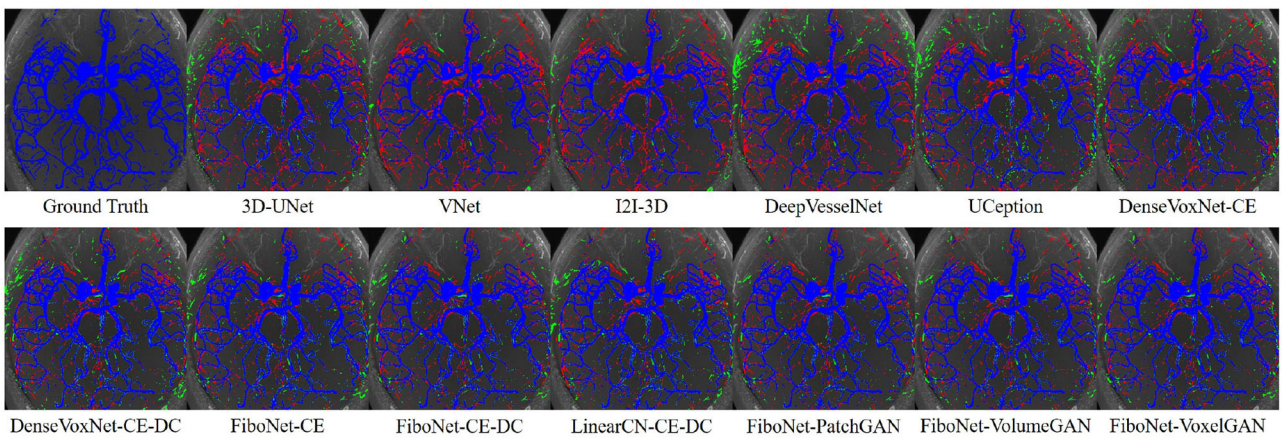


FIGURE 4 | Maximum intensity projections of each method. Pixels in blue represent true positives. Pixels in green represent false positives. Pixels in red represent false negatives.

- Symmetric Average Surface Distance (*SASD*) Yeghiazaryan and Voiculescu (2015):
  $SASD = (mean(Dist(P, G)) + mean(Dist(G, P)))/2$
- Symmetric 95% Hausdorff Distance (*SHD*95) Dubuisson and Jain (1994):
  $SHD95 = (P_{95}(Dist(P, G)) + P_{95}(Dist(G, P)))/2$

where P represents the predicted segmentation maps of each network, G represents the annotated ground truth, $Dist(P, G)$ is the set of distances from boundary voxels of P, δ_P, to the nearest boundary voxel in δ_G, i.e.,:

$$Dist(P, G) = \{\min_{x \in \delta_P} \|x - y\|^2 | y \in \delta_G\} \quad (9)$$

wherein $P_{95}(Dist(\cdot))$ is the 95th percentile of $Dist(\cdot)$.

Among these metrics, the *DC* is a commonly used metric for measuring the general agreement between the generated segmentation maps and the annotated ground truth. It can be utilized to evaluate the general performance of each method. *SASD* reflects the globalized average boundary agreement, while *SHD*95 is usually applied to evaluate the localized disagreement.

4.3. Qualitative Results

Maximum intensity projection, along with slice-wise result, of each method was shown in **Figures 4**, **5**, respectively, wherein blue, red, and green pixels represent true positives, false negatives, and false positives, respectively.

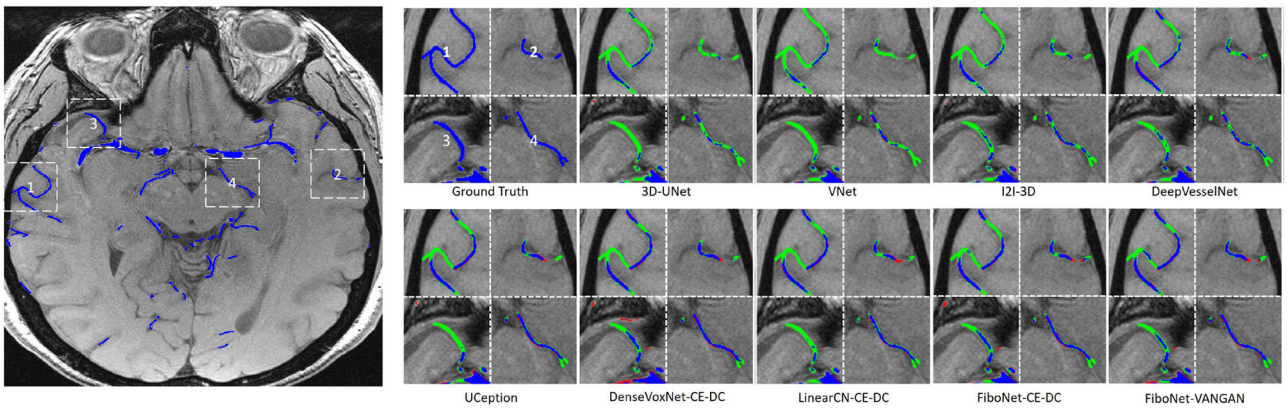


FIGURE 5 | Visual results of the four different white dash-line rectangles of each method. Pixels in blue represent true positives. Pixels in green represent false positives. Pixels in red represent false negatives.

TABLE 4 | *P*-value, Cohen's D, and intraclass correlation coefficient (ICC) values between the proposed method, FiboNet-VANGAN, and each of its counterparts.

	***P*-value/Cohen's D / ICC**	
Methods	**Healthy dataset**	**Brain atrophy dataset**
3D-UNet	$\mathbf{7.6}e^{-7}$ / **3.26** / **0.84**	$1.7e^{-4}$ / **4.89** / **0.93**
V-Net	$8.9e^{-7}$ / **3.03** / **0.82**	$6.3e^{-7}$ / **6.94** / **0.96**
I2I-3D	$1.9e^{-7}$ / **3.15** / **0.83**	$2.0e^{-6}$ / **6.17** / **0.95**
DeepVesselNet	$2.1e^{-7}$ / **2.86** / **0.80**	$7.3e^{-7}$ / **5.90** / **0.95**
Uception	$1.9e^{-6}$ / **1.87** / **0.83**	$2.6e^{-6}$ / **8.25** / **0.97**
DenseVoxNet-CE	$1.3e^{-6}$ / **2.19** / **0.70**	$5.7e^{-7}$ / **7.66** / **0.97**
DenseVoxNet-CE-DC	$2.7e^{-8}$ / **1.22** / **0.42**	$4.5e^{-7}$ / **2.25** / **0.73**
FiboNet-CE	$2.1e^{-4}$ / **0.81** / **0.24**	$4.0e^{-4}$ / **4.04** / **0.90**
FiboNet-CE-DC	$1.3e^{-6}$ / **0.35** / **0.06**	$4.2e^{-1}$ / 0.01 / 0.02
LinearCN-CE-DC	$1.3e^{-5}$ / **0.64** / **0.16**	$6.1e^{-6}$ / **1.51** / **0.55**
FiboNet-PatchGAN	$2.9e^{-1}$ / 0.04 / 0.00	$2.8e^{-1}$ / 0.28 / 0.04
FiboNet-VolumeGAN	$1.8e^{-3}$ / **0.21** / **0.01**	$6.2e^{-2}$ / **0.28** / **0.04**

Bold values indicate improvements of statistical significance.

4.4. Quantitative Results

In **Table 3**, we listed the quantitative comparisons of different methods. Aside from the results on the rest 6 testing volumes from the healthy dataset, we also include the results for the 9 testing volumes from the brain atrophy dataset.

4.5. Statistical Analysis

Paired-sample t-test is adopted to evaluate the significance of the improvement of the average DC by the proposed method. We contrast the proposed FiboNet-VANGAN with other methods listed in **Table 3**. One-tailed results are reported since we only care about the significance of the improvements. Therefore, our null and alternative hypothesis should format separately as:

$$H_0 : \mu_1 = \mu_2 \tag{10}$$

$$H_{\alpha=0.01} : \mu_1 > \mu_2 \tag{11}$$

where H_0 represents the null hypothesis and H_α alternative hypothesis. μ_1 represents the average DC value from the proposed method, FiboNet-VANGAN, and μ_2 the contrasting method accordingly. In this hypothesis testing, we set the significance-level to 0.01. Here, we report the test results of both datasets. In the healthy dataset, $\mu_i, (i = 1, 2)$ is the average DC value over the 12 volumes. In the brain atrophy dataset, $\mu_i, (i = 1, 2)$ is the average DC value over the 9 volumes. Aside from p-values, we also included values of Cohen's D and intraclass correlation coefficient (ICC) to test the robustness of the proposed model. The test results are listed in **Table 4**, wherein p-values showing statistical significance, along with the corresponding values of Cohen's D and ICC, were represented by black bold fonts.

In the healthy dataset, the proposed method significantly outperforms its counterparts, demonstrating a significant improvement of DC value by our method. In the brain atrophy dataset, significant improvement can also be observed except

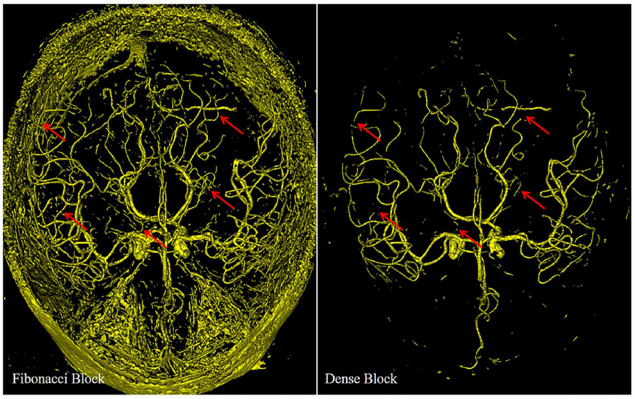


FIGURE 6 | Different results from the deep supervision branch of FiboNet and DenseVoxNet, illustrating the different detection performance of Fibonacci block and the dense block. Red arrows indicate regions where FiboNet could extract more complete vasculature than DenseVoxNet.

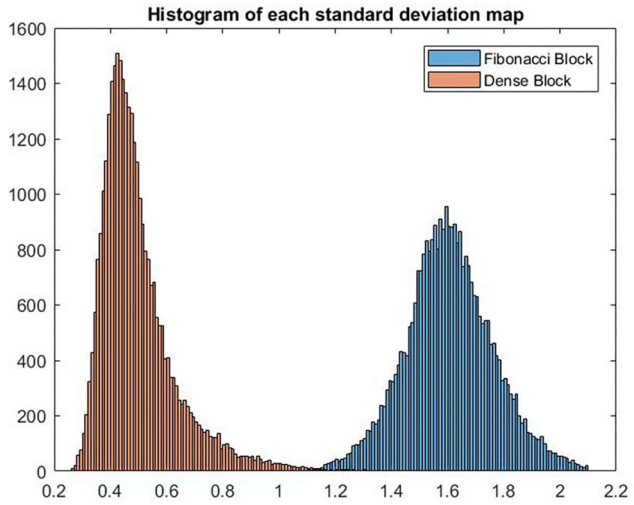


FIGURE 7 | VHistograms of the two standard deviation maps calculated using the feature maps from Fibonacci block and dense block, respectively.

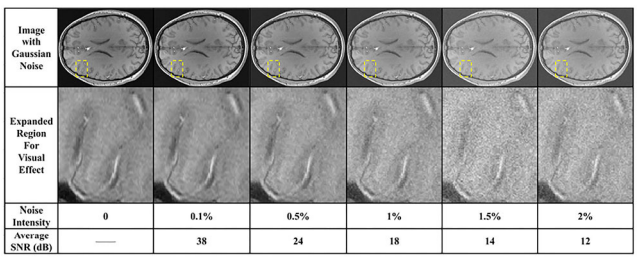


FIGURE 8 | Illustration of a slice with different magnitudes of Gaussian noise. Window level is 240 and window width is 240. The region marked by yellow dash-line rectangle is expanded for better visual effects. Averaged : signal to noise ratioes (SNRs) between images with added noise and the original image were listed in the last row.

the result from FiboNet-PatchGAN. Indeed, FiboNet-PatchGAN can achieve comparable accuracy in the brain atrophy dataset with the proposed method, but it underperforms in the healthy dataset.

4.6. Evaluation of the Diversity of Feature Maps

The feature map generated by the Fibonacci block (or the dense block) is a matrix of shape $[H, W, D, C]$, where H, W, D, C represents the value of height, width, depth, and channel, respectively. By calculating the standard deviation along the fourth dimension, we can get the standard deviation map, shaping like $[H, W, D]$, of the channels. Based on the channel-wise standard deviation map, histograms are calculated to compare the diversity of feature maps generated by these two blocks.

In **Figure 6**, we display the output of the deep supervision branch to validate the performance of the Fibonacci block and the dense block.

Compared with the results of the dense block, more candidate voxels are extracted by the Fibonacci block. Few vascular voxels are left out by the Fibonacci block. While in the result of the dense block, though the vascular regions compose the major part of the result, distal vascular regions and many small vascular branches are missing. Actually, both results stem from minimizing the loss function. If features of distal vascular regions are learned by the dense block, these voxels will be segmented so as to reduce the loss value. But in the dense block, due to its regularizing effect, these features are probably not learned. The Fibonacci block derives from the dense block by weakening its regularizing effect, as described in section II, which leads to the improvement of the detection ability of network. This results in the segmentation of distal vascular voxels but also many background voxels. If features within these regions are not learned, there is no chance that they can be segmented. Therefore, we can conclude from this observation that diverse feature maps are generated by the Fibonacci block, that is, the FiboNet.

In **Figure 7**, we show the histograms of channel-wise standard deviation map calculated using the feature maps generated by Fibonacci block and dense block. It can be concluded from the histograms that standard deviations of the feature maps generated by the Fibonacci block are larger than their counterparts from the dense block, which means the Fibonacci block can generate more diversified feature maps than the dense block.

4.7. Noise Robustness

Noise is commonly observed in medical imaging. Automatic methods should be robust to different magnitudes of noise. In order to evaluate the robustness of different methods, we apply the different methods on test set with different magnitudes of noise. Gaussian noise with 0.1%, 0.5%, 1%, 1.5%, and 2% of the image magnitude is added. In **Figure 8**, we demonstrate

a slice with no Gaussian noise and the slices with different magnitudes of Gaussian noise. A rectangle region marked by the yellow dash-line is expanded for better visual effects. The increasing magnitudes of Gaussian noise deteriorate the contrast of the original slice. Some vascular boundaries are blurred and degenerated, thus hard to be distinguished from background egions.

In **Figure 9**, we display the quantitative comparisons of noise robustness of each method. We use the vertical axis to represent the drop-to performance calculated via:

$$DTP = \frac{DC_p}{DC_0} \times 100\% \tag{12}$$

where DC_p represents the DC value obtained from image with p% noise, and DC_0 is the DC value when no Gaussian noise is added. The horizontal axis represents the different magnitudes of Gaussian noise added. It can be observed that the performance of each method drops with the increasing strength of noise. But the robustness of each method is quite different.

Among these networks, 3D-UNet turns out to be the least robust to noise. In this network, batch normalization is not included, which will lead to ineffective training. Dropout layer to enhance the robustness of network is not included either. An improved robustness can be observed in V-Net where residual connection of feature maps is utilized in network design. Residual connection helps promote the effective training of networks. Therefore, better robustness can be realized. However, in the implementation of V-Net, residual connection is incorporated in the first layer of each scale of feature space. No input transition layer (Smith and Topin, 2016) is used. Usually, an input transition layer is utilized for two reasons. One is to transform the input from raw image space to its corresponding feature space. The other reason is to increase the number of channels of input image, which allows the input data to be examined in many ways. This type of layer has been adopted in many successful networks (Krizhevsky et al., 2012; Szegedy et al., 2015, 2016, 2017; He et al., 2016). But in V-Net, a residual block is directly adopted after the input image. Noise from raw image is propagated to and further influence deep layers. Hence, only limited improvement is achieved. Better situations can be observed in I2I-3D, DenseVoxNet, FiboNet, and LinearCN. Within these networks, convolution layers are utilized as input transition layers to transform from raw image space to feature space. Influence of noise can be reduced by the nonlinear transformation of input transition layers. Attentions should also be paid to the improvement brought about by incorporating DC as one of the loss functions. In DenseVoxNet, very limited improvement of noise robustness is achieved by DC loss. While in FiboNet, the improvement of noise robustness is very obvious and much larger than that in DenseVoxNet. Finally, PatchGAN and VANGAN help improve the noise robustness by a small margin. Therefore, in practice we should trade off between the segmentation accuracy and noise robustness when designing network architectures.

4.8. Size of Training Patch

In section II, we elaborate the reason to utilize patch-wise training strategy. However, in the original implementation of U-Net (Ronneberger et al., 2015), image of size 572 × 572 is applied for training and testing. In the original implementation of V-Net (Milletari et al., 2016), image of size 128 × 128 × 64 is adopted. Due to the limited graphic memory, we cannot carry out the experiment using images of size 572 × 572 × 572 for 3D-UNet.

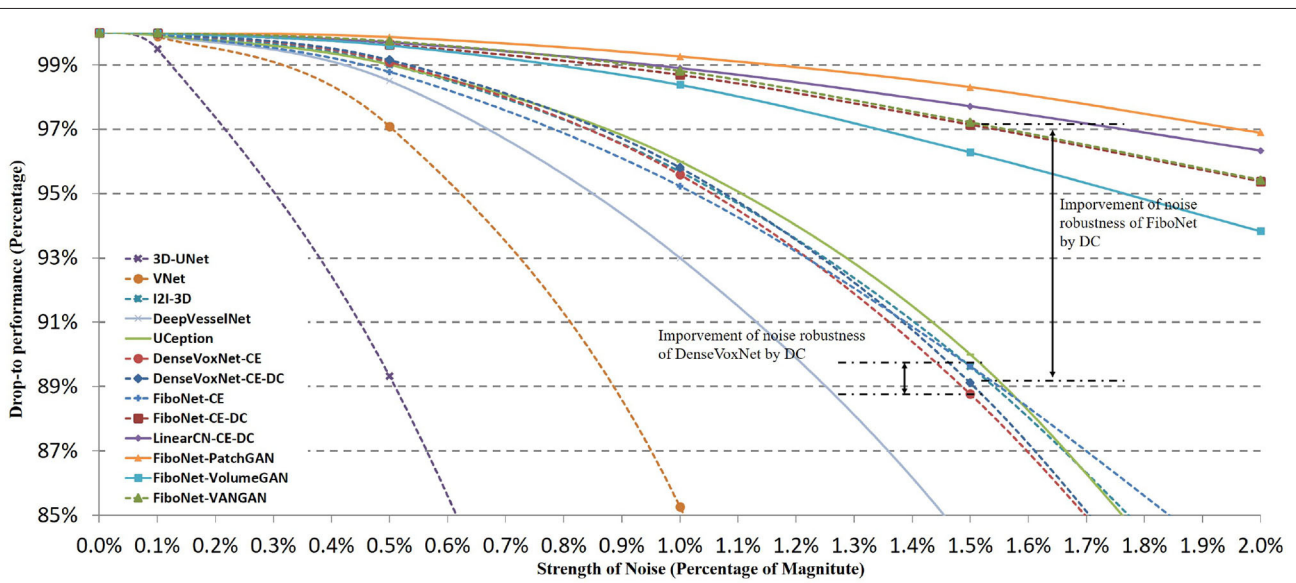


FIGURE 9 | Influences of different magnitudes of Gaussian noise on the segmentation accuracy of different methods. Horizontal axis represents the strength of noise added. Vertical axis represents drop-to performance of a certain method compared with the situation when no Gaussian noise is added.

But we try the experiments with patches of size $128 \times 128 \times 64$ and $64 \times 64 \times 64$ on V-Net, LinearCN, and FiboNet. The same data splitting strategy of the healthy dataset as described in section IV.A is used. These three networks are trained for 40 epochs, which means it goes over the training dataset for 40 times either by the way when patch size of $128 \times 128 \times 64$ is adopted or the way patch size of $64 \times 64 \times 64$ is adopted.

The results are summarized in **Table 5**. It turns out that better results are obtained using patches of size $64 \times 64 \times 64$, which in turn supports our assumption that small patches can be regarded as an implicit data augmentation for cerebrovascular segmentation due to potential similar feature from small patches.

TABLE 5 | Results using different size of training patches.

Healthy dataset			
Method / Patch Size	DC		
	V-Net	LinearCN	FiboNet
$64 \times 64 \times 64$	0.7191	0.7927	0.7967
$128 \times 128 \times 64$	0.6636	0.7125	0.7748
ΔDC	0.0555	0.0802	0.0219

5. DISCUSSION

Clinically acceptable segmentation accuracy has yet to be defined for assisting surgical planning, the preventive diagnosis and quantitative analysis of cerebral vascular diseases. However, following three aspects could be the focuses for developing algorithms of cerebrovascular segmentation:

- High segmentation accuracy of vasculature—This is important for many clinical applications like invasive surgical planning, preventive diagnosis, etc. In invasive surgical planning, we should include accurate surface model to plan a proper path for intervention. In preventive diagnosis, clinicians can make a comprehensive and meaningful evaluation based on geometric features of vasculature, which should also base on high segmentation accuracy.
- Detection of vessel with small radius—Some aneurysms occur at the boundaries of small vessels. Failing to detect these vessels may lead to mis-evaluate the risk of strokes.
- Robustness to noise—Noise is an inevitable phenomenon during imaging. If automatic algorithms are not robust enough to noise, the segmentation performance will drop. This leads to inaccurate vascular information, based on which it is hard to carry out meaningful postprocessing like surgical planning, geometric analysis of vessels, etc.

In this paper, we present a deep learning-based algorithm to segment the cerebral vasculature for TOF-MRA images. Compared with the counterparts trained using CE as the only loss (DenseVoxNet-CE and FiboNet-CE in **Table 3**), the inclusion of *DC* improves the segmentation result by $\sim$ 2%. The result of DenseNet-CE-DC is close but still lower than our FiboNet-CE, which also validates the effectiveness of FiboNet. In order to alleviate the large variations in vascular anatomies and voxel intensities, we propose a VAN. VAN is adopted as the regularizer for voxel-wise distribution consistency between the predictions and ground truth by incorporating it in the adversarial training. To mitigate the influence of class imbalance on the convergence of D, FL is utilized as $D's$ loss function. To relieve the influence of class imbalance on the convergence of G, the FiboNet, we propose to use the addition of cross-entropy and DC as $G's$ loss function. In this paper, we also propose the Fibonacci connection of feature maps and incorporate it in our FiboNet. Compared with the dense connection, this type of feature aggregation is suitable to be placed in shallow layers for diversified feature maps. Noise robustness is improved by a large margin by incorporating the DC in the training of FiboNet.

Comparatively, 3D-UNet turns out to demonstrating the weakest performance on both datasets. Maybe, only low level feature like voxel intensity is learned by 3D-UNet. From the results of V-Net and I2I-3D, conclusion can be made that, deep supervision helps improve its performance on abnormal dataset due to its intrinsic property to promote highly discriminative feature maps between convolutional layers (Lee et al., 2015), which also coincides with our motivation of Fibonacci connection. Comparatively, cross-hair filters used in DeepVesselNet helps achieve comparable segmentation accuracy but with improved memory usage and thus faster training speed.

Refer to the results of DenseVoxNet-CE-DC, FiboNet-CE-DC, and LinearCN-CE-DC for comparison of different connection types. Within these results, the DenseVoxNet-CE-DC gets the smallest *DC* value. This is probably due to the regularizing problem. The performance of LinearCN-CE-DC is a little weaker but very close to the performance of FiboNet-CE-DC. From **Figure 4**, we can find out that LinearCN-CE-DC performs better in vascular regions while it does worse in background regions. Worse local agreement is achieved by LinearCN-CE-DC than FiboNet-CE-DC. This can also be concluded from its value of *SHD*95, which is much larger than FiboNet-CE-DC. On the contrary, the FiboNet structure applies the feature-reusing strategy by concatenating the output from previous two layers as the input of the current layer. Compared with LinearCN, this type of connection encourages better back propagation of gradients. Compared with dense connection, this type of connection encourages diversified feature maps, which is better for vascular segmentation due to all kinds of vascular features.

The incorporation of adversarial training helps improve the segmentation performance on both datasets. The improvements from VolumeGAN is slight on both datasets. Comparatively, PatchGAN can achieve much higher improvement than VolumeGAN on the brain atrophy dataset. This may imply that small size of training patches are preferred in deep learning-based vascular segmentation since vasculatures are likely to share similar feature from a small patch instead of a large one (refer to section IV.G for details). VolumeGAN is deeper than PatchGAN, making it harder to train. Similar phenomenon that PatchGAN outperforms VolumeGAN is observed in Isola et al. (2017). Also, when networks go deeper, fewer voxels are output by VolumeGAN, making it ineffective in producing

stable and sufficient gradient feedback for network training (Xue et al., 2018). The application of voxel-wise adversarial training (VANGAN) can relieve above problems. VANGAN has greater depth than VolumeGAN and the output size of VANGAN is the same as its input, ensuring sufficient backpropagation of gradient. Also, best and more consistent performance is realized on both datasets using VANGAN than that of PatchGAN. This can be concluded by the largest *DC* value and the smallest standard deviation of VANGAN among the methods.

Though improvements have been made, imperfections of our method do exist. First of all, our method only deals with the segmentation of cerebral vasculature. Centerline of the extracted vasculature cannot be detected explicitly. In applications involving the geometric information of vessels, centerlines are very important. With centerlines, the accuracy of the automatically reconstructed topology can also be measured. Second, we observe some false merges in the segmentation of vessels that are quite close to each other. When vessels are quite close to each other, only a blurry small gap of two to three voxels is supposed to separate the vessels. In the situation like this, existing methods including our FiboNet fail to separate the vessels. Instead, the blurry small gap is also recognized as vascular region, leading to false merge of vessels. Third, in patch-wise training, space information is lost for now. Networks cannot perceive the global space position of each voxel within each subject. Therefore, vessels outside the brain will also be extracted. In the future, attentions will paid to addressing above problems.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

BG, BL, and XB conceived the project. BG implemented it and wrote the draft. BG, BL, XB, and FZ edited it. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the National Key R&D Program of China under Grant No. 2018YFA0704101, and the National Natural Science Foundation of China (61601012, U1736217).

REFERENCES

Brisman, J. L., Song, J. K., and Newell, D. W. (2006). Cerebral aneurysms. *N Engl. J. Med*. 355, 928–939. doi: 10.1056/NEJMra052760

Buda, M., Maki, A., and Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw*. 106, 249–259. doi: 10.1016/j.neunet.2018.07.011

Chen, L., Xie, Y., Sun, J., Balu, N., Mossa-Basha, M., Pimentel, K., et al. (2017a). Y-net: 3d intracranial artery segmentation using a convolutional autoencoder. *arXiv:1712.07194*. doi: 10.1109/BIBM.2017.8217741

Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., and Feng, J. (2017b). "Dual path networks," in *Advances in Neural Information Processing Systems*, 4467–4475.

Christ, P. F., Ettlinger, F., Grün, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S., et al. (2017). Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. *arXiv:1702.05970*.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302. doi: 10.2307/1932409

Dolz, J., Ayed, I. B., Yuan, J., and Desrosiers, C. (2018). "Isointense infant brain segmentation with a hyper-dense connected convolutional neural network," in *15th International Symposium on Biomedical Imaging* (Washington, DC: IEEE), 616–620.

Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., and Heng, P. (2016). "3D deeply supervised network for automatic liver segmentation from CT volumes," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer), 149–157.

Dubuisson, M.-P., and Jain, A. K. (1994). "A modified hausdorff distance for object matching," in *Proceedings of 12th International Conference on Pattern Recognition* (Jerusalem: IEEE), 566–568.

Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., et al. (2018). Automatic multi-organ segmentation on abdominal ct with dense v-networks. *IEEE Trans. Med. Imaging* 37, 1822–1834. doi: 10.1109/TMI.2018.2806309

Goceri, E., Shah, Z. K., and Gurcan, M. N. (2017). Vessel segmentation from abdominal magnetic resonance images: adaptive and reconstructive approach. *Int. J. Numer. Method Biomed. Eng*. 33:e2811. doi: 10.1002/cnm.2811

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (Montréal), 2672–2680.

Han, Z., Wei, B., Mercado, A., Leung, S., and Li, S. (2018). Spine-gan: Semantic segmentation of multiple spinal structures. *Med. Image Anal*. 50:23–35. doi: 10.1016/j.media.2018.08.005

Hassouna, M. S., Farag, A. A., Hushek, S., and Moriarty, T. (2006). Cerebrovascular segmentation from tof using stochastic models. *Med. Image Anal*. 10, 2–18. doi: 10.1016/j.media.2004.11.009

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* (Las Vegas, NV), 770–778.

Huang, G., Liu, Z., Weinberger, K. Q., and van der Maaten, L. (2017). Densely connected convolutional networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recogn*. 1, 3. doi: 10.1109/CVPR.2017.243

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning* (Lille), 448–456.

Iqbal, S. (2013). A comprehensive study of the anatomical variations of the circle of willis in adult human brains. *J. Clin. Diagn. Res*. 7, 2423. doi: 10.7860/JCDR/2013/6580.3563

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu: IEEE), 5967–5976.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, Harrahs and Harveys, 1097–1105.

Kushibar, K., Valverde, S., González-Villà, S., Bernal, J., Cabezas, M., Oliver, A., et al. (2018). Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features. *Med. Image Anal*. 48, 177–186. doi: 10.1016/j.media.2018.06.006

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., and Tu, Z. (2015). "Deeply-supervised nets," in *Artificial Intelligence and Statistics* (San Diego, CA), 562–570.

Lesage, D., Angelini, E. D., Bloch, I., and Funka-Lea, G. (2009). A review of 3d vessel lumen segmentation techniques: models, features and extraction schemes. *Med. Image Anal*. 13, 819–845. doi: 10.1016/j.media.2009.07.011

Li, Y., Qiao, Z., Zhang, S., Wu, Z., Mao, X., Kou, J., et al. (2017). A novel method for low-contrast and high-noise vessel segmentation and location in venipuncture. *IEEE Trans. Med. Imaging* 36, 2216–2227. doi: 10.1109/TMI.2017.2732481

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 2980–2988.

Luc, P., Couprie, C., Chintala, S., and Verbeek, J. (2016). "Semantic segmentation using adversarial networks," in *NIPS Workshop on Adversarial Training* (Barcelona).

Lugauer, F., Zheng, Y., Hornegger, J., and Kelm, B. M. (2014). "Precise lumen segmentation in coronary computed tomography angiography," in *International MICCAI Workshop on Medical Computer Vision* (Lugauer: Springer), 137–147.

Mapayi, T., Tapamo, J.-R., and Viriri, S. (2015). Retinal vessel segmentation: a comparative study of fuzzy c-means and sum entropy information on phase congruency. *Int. J. Adv. Rob. Syst.* 12, 133. doi: 10.5772/60581

McKinney, A., Palmer, C., Truwit, C., Karagulle, A., and Teksam, M. (2008). Detection of aneurysms by 64-section multidetector ct angiography in patients acutely suspected of having an intracranial aneurysm and comparison with digital subtraction and 3d rotational angiography. *Am. J. Neuroradiol.* 29, 594–602. doi: 10.3174/ajnr.A0848

Merkow, J., Marsden, A., Kriegman, D., and Tu, Z. (2016). "Dense volume-to-volume vascular boundary detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Athens: Springer), 371–379.

Millán, R. D., Dempere-Marco, L., Pozo, J. M., Cebral, J. R., and Frangi, A. F. (2007). Morphological characterization of intracranial aneurysms using 3-d moment invariants. *IEEE Trans. Med. Imaging* 26, 1270–1282. doi: 10.1109/TMI.2007.901008

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Fourth International Conference on 3D Vision* (Stanford, CA: IEEE), 565–571.

Nain, D., Yezzi, A., and Turk, G. (2004). "Vessel segmentation using a shape driven flow," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Brittany: Springer), 51–59.

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, 807–814.

Nowinski, W. L., Volkau, I., Marchenko, Y., Thirunavuukarasuu, A., Ng, T. T., and Runge, V. M. (2009). A 3d model of human cerebrovasculature derived from 3t magnetic resonance angiography. *Neuroinformatics* 7, 23–36. doi: 10.1007/s12021-008-9028-8

Oliveira, D. A., Feitosa, R. Q., and Correia, M. M. (2011). Segmentation of liver, its vessels and lesions from ct images for surgical planning. *Biomed. Eng. Online* 10, 30. doi: 10.1186/1475-925X-10-30

Raghavan, M. L., Ma, B., and Harbaugh, R. E. (2005). Quantified aneurysm shape and rupture risk. *J. Neurosurg.* 102, 355–362. doi: 10.3171/jns.2005.102.2.0355

Rätsep, M., Paolozza, A., Hickman, A., Maser, B., Kay, V., Mohammad, S., et al. (2016). Brain structural and vascular anatomy is altered in offspring of pre-eclamptic pregnancies: a pilot study. *Am. J. Neuroradiol.* 37, 939–945. doi: 10.3174/ajnr.A4640

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 234–241.

Sanchesa, P., Meyer, C., Vigon, V., and Naegel, B. (2019). "Cerebrovascular network segmentation of mra images with deep learning," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (Venice: IEEE), 768–771.

Schaap, M., Metz, C. T., van Walsum, T., van der Giessen, A. G., Weustink, A. C., Mollet, N. R., et al. (2009). Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Med. Image Anal.* 13, 701–714. doi: 10.1016/j.media.2009.06.003

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. doi: 10.1109/TPAMI.2016.2572683

Smith, L. N., and Topin, N. (2016). Deep convolutional neural network design patterns. *arXiv:1611.00847*.

Soares, J. V., Leandro, J. J., Cesar, R. M., Jelinek, H. F., and Cree, M. J. (2006). Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging* 25, 1214–1222. doi: 10.1109/TMI.2006.879967

Sønderby, C. K., Caballero, J., Theis, L., Shi, W., and Huszár, F. (2016). Amortised ma inference for image super-resolution. *arXiv:1610.04490*.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, CA).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA: IEEE).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 2818–2826.

Tetteh, G., Efremov, V., Forkert, N. D., Schneider, M., Kirschke, J., Weber, B., et al. (2018). Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *arXiv:1803.09340*.

Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., et al. (2018). Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Trans. Med. Imaging* 37, 1562–1573. doi: 10.1109/TMI.2018.2791721

Xie, S., and Tu, Z. (2015). "Holistically-nested edge detection," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago: IEEE), 1395–1403.

Xue, Y., Xu, T., Zhang, H., Long, L. R., and Huang, X. (2018). Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. *Neuroinformatics* 16, 383–392. doi: 10.1007/s12021-018-9377-x

Yeghiazaryan, V., and Voiculescu, I. (2015). *An overview of current evaluation methods used in medical image segmentation*. Technical report, Tech. Rep. CS-RR-15-08, Department of Computer Science, University of Oxford, Oxford, UK.

Yu, L., Cheng, J., Dou, Q., Yang, X., Chen, H., Qin, J., et al. (2017). "Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer), 287–295. doi: 10.1007/978-3-319-68520-5

Zarrinkoob, L., Ambarki, K., Wåhlin, A., Birgander, R., Eklund, A., and Malm, J. (2015). Blood flow distribution in cerebral arteries. *J. Cereb. Blood Flow Metab.* 35, 648–654. doi: 10.1038/jcbfm.2014.241

Zheng, Y. (2016). "Model-based 3d cardiac image segmentation with marginal space learning," in *Medical Image Recognition, Segmentation and Parsing* (Elsevier), 383–403.

Zheng, Y., Shen, J., Tek, H., and Funka-Lea, G. (2012). "Model-driven centerline extraction for severely occluded major coronary arteries," in *International Workshop on Machine Learning in Medical Imaging* (Nice: Springer), 10–18.

Zheng, Y., Tek, H., and Funka-Lea, G. (2013). "Robust and accurate coronary artery centerline extraction in cta by combining model-driven and data-driven approaches," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Nagoya: Springer), 74–81.

Zheng, Y., Yang, D., John, M., and Comaniciu, D. (2014). Multi-part modeling and segmentation of left atrium in c-arm ct for image-guided ablation of atrial fibrillation. *IEEE Trans. Med. Imaging* 33, 318–331. doi: 10.1109/TMI.2013.2284382

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

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

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

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 17 November 2021
doi: 10.3389/fnins.2021.774857

Electroencephalogram-Based Motor Imagery Classification Using Deep Residual Convolutional Networks

Jing-Shan Huang[1,2], Wan-Shan Liu[1,2], Bin Yao[1,2], Zhan-Xiang Wang[3,4,5], Si-Fang Chen[5]* and Wei-Fang Sun[6]**

[1] School of Aerospace Engineering, Xiamen University, Xiamen, China, [2] Shenzhen Research Institute of Xiamen University, Shenzhen, China, [3] Institute of Neurosurgery, School of Medicine, Xiamen University, Xiamen, China, [4] Xiamen Key Laboratory of Brain Center, The First Affiliated Hospital of Xiamen University, Xiamen, China, [5] Department of Neurosurgery, The First Affiliated Hospital of Xiamen University, Xiamen, China, [6] College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Wei Cao,
Huaqiao University, China
Zhiguo Wan,
Xi'an Shiyou University, China
Wenrong Xiao,
China Three Gorges University, China

***Correspondence:**
Bin Yao
aeroiiet@126.com
Si-Fang Chen
csfsong143@aliyun.com
Wei-Fang Sun
swf@wzu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 13 September 2021
Accepted: 22 October 2021
Published: 17 November 2021

Citation:
Huang J-S, Liu W-S, Yao B, Wang Z-X, Chen S-F and Sun W-F (2021) Electroencephalogram-Based Motor Imagery Classification Using Deep Residual Convolutional Networks. Front. Neurosci. 15:774857. doi: 10.3389/fnins.2021.774857

The classification of electroencephalogram (EEG) signals is of significant importance in brain-computer interface (BCI) systems. Aiming to achieve intelligent classification of motor imagery EEG types with high accuracy, a classification methodology using the wavelet packet decomposition (WPD) and the proposed deep residual convolutional networks (DRes-CNN) is proposed. Firstly, EEG waveforms are segmented into sub-signals. Then the EEG signal features are obtained through the WPD algorithm, and some selected wavelet coefficients are retained and reconstructed into EEG signals in their respective frequency bands. Subsequently, the reconstructed EEG signals were utilized as input of the proposed deep residual convolutional networks to classify EEG signals. Finally, EEG types of motor imagination are classified by the DRes-CNN classifier intelligently. The datasets from BCI Competition were used to test the performance of the proposed deep learning classifier. Classification experiments show that the average recognition accuracy of this method reaches 98.76%. The proposed method can be further applied to the BCI system of motor imagination control.

Keywords: electroencephalogram (EEG), motor imagery (MI), wavelet packet decomposition (WPD), residual, convolutional neural networks

INTRODUCTION

Electroencephalogram (EEG) is a common biological signal in the medical field. People obtain EEG signals by collecting and recording the potential changes of the superficial skin of the head, and characterize the activity characteristics of the brain. The research of EEG signals is widely used in various aspects. In the field of biomedicine, EEG signal researches help doctors diagnose neurological diseases, such as frostbite, epilepsy, Alzheimer's disease, childhood developmental disorders, schizophrenia, Parkinson's disease and other functional diseases. EEG can detect sleep quality, fatigue driving, and drunk driving. It can also study human brain functions such as emotion, cognition, memory, and sports. At present, a mainstream research direction is to use brain waves to control objects. Through certain processing of EEG signals, certain parts of the body can be controlled to make certain actions. At present, certain research results have been achieved.

For example, people invented instruments and tools to facilitate the lives of the disabled. Motor imaging signal is a kind of brain electrical signal, which is often used in brain-computer interface (BCI).

Brain-computer interface technology is a system that does not pass through the normal physiological pathways of the human body, but allows the brain to directly transmit information or control commands to computers or related instruments (Gong, 2014). When the brain performs motor imagination, the corresponding brain regional potential will also change accordingly due to the different imagination of the imagination activity. The corresponding changes in the brain are detected, and the computer is used to convert these detected change signals into instructions to control the lower computer (Cilliers and Van Der Kouwe, 1993). The EEG signal is an important part of the BCI system. The acquisition of motor imagery EEG signals is the first step to realize the operation of the BCI system. Then the EEG signal is processed and decoded. Finally, the EEG signal is translated into "machine language" through the control instruction conversion module to drive the external equipment, so that the purpose of human-computer interaction can be realized.

Pattern recognition of various states of the human body based on brainwave detection is a very popular research topic, and it has produced quite constructive results in many fields. K. Polat proposed a classification method of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform (Kemal and Salih, 2007). M. V. M. Yeo and X. P. Li used support vector machines for pattern recognition and developed a method to automatically detect the driver's fatigue driving state (Mervyn et al., 2009). The method recognition accuracy reached 99.3%, and it can reliably predict the transition from alertness to drowsiness. T. Nguyen proposed a threshold method to identify blinking state, and achieved good results in the detection results (Nguyen et al., 2013). Wang proposed an EEG eye state identification method using incremental attribute learning with time-series classification, and the method finally achieved an accuracy rate higher than average (Wang et al., 2014). S. K. Satapathy used neural network and support vector machine to perform brain wave-based pattern recognition for epilepsy and has achieved good recognition results (Satapathy et al., 2017). G. Anumanchipalli uses the RNN deep learning model to directly read the thoughts in the brains of paralyzed patients using a BCI. The spoken sentences can reach 150 words per minute, which is close to the normal level of people (Anumanchipalli et al., 2019).

Since the end of the 1960s, humans have studied BCI technology for more than 50 years. Since the 21st century, the research of BCI has become more and more prosperous. Four international BCI competitions were successfully held. Researchers have systematically analyzed the processing methods of EEG signals and produced some mature applications. The BCI laboratory team at the Cognitive Institute of Graz University of Technology in Austria first implemented a BCI based on online EEG classification. The team developed a variety of BCI systems using motor imaging brain electrical signals, including imagining different limb movements to control the movement of the wheelchair (Pfurtscheller et al., 1993, 2006), and using brain waves to control the movement of the mouse to find coins (Pfurtscheller et al., 2000). The BCI Research Institute in Berlin has developed a typing system (Benjamin et al., 2003). The subjects selected different characters for typing by imagining the movements of the left hand, right hand, and foot. The Washington Research Center in the United States uses different EEG rhythm signals generated by motor imagination to realize the free movement of the virtual cursor in three-dimensional space (McFarland et al., 2010). The BSI-TOYOTA Collaboration Center in Japan has successfully developed a real-time control wheelchair using brain waves. By imagining the front, left, and right to control the direction of the wheelchair, a 125 ms response control system for the electric wheelchair can be realized (Bai, 2010). Gao Shangkai of Tsinghua University used the characteristics of motor imaginary EEG signals to develop a system that uses EEG signals to control robot dogs playing football (Wang et al., 2007). Xu Baoguo of Southeast University controlled the robot arm to make corresponding actions based on imagining the movement of the hand (Xu et al., 2011). The average accuracy of motor imaging EEG for manipulator control is 88%. Li Yuanqing of South China University of Technology designed a hybrid BCI system that combines motor-imaging EEG signals and P300 signals, which control the horizontal and vertical movement of the cursor, respectively (Li et al., 2010).

Although many laboratory results have been achieved in the research of BCI technology, there are still few products that can be applied in real life. The BCI technology is still in the stage of theoretical research and laboratory development, and the application system needs to be further improved. There are still many key technologies that need to be improved. Firstly, the existing BCI systems have poor adaptability. Different individuals have different physiological functions, so people's EEG physiological responses to the same task will also be different. When the same individual performs the same motor imagination activity in different mental states, the EEG response may also be different. The adaptability of the future BCI system should not only meet the differences between different individuals, but also meet the changes of different states of the same individual. Secondly, the recognition speed and accuracy of the BCI system need to be improved. Classification accuracy and recognition speed are the most commonly performance evaluation indicators used in BCI system. The existing feature extraction methods used for classification of motion imaging tasks are relatively complicated. The large amount of data leads to long calculation time and slow system processing speed. However, reducing the amount of data or simplifying the signal feature extraction method will cause the classification accuracy to decrease. The key challenge of current research is to speed up the processing speed of the system while ensuring the accuracy of classification. In addition, it is relatively difficult to integrate technologies in different fields with BCI system. Promoting the technical integration of BCI application systems and being accepted by users is also an important practical problem faced by BCI systems.

In this article, we propose an accurate EEG signal classification method using Deep Residual Convolutional Neural Network (DRes-CNN). The EEG signals in BCI Competition 2005 data

set IVa and BCI Competition 2003 data set III are selected as the original data. The wavelet packet decomposition (WPD) was used for preprocessing to obtain the characteristics of EEG signals. Subsequently, the reconstructed EEG signals of different frequency bands were used as the input of DRes-CNN to finally identify and classify the EEG types. The classification results show that the average accuracy of the proposed DRes-CNN model can reach 98.76%. The rest of this article is organized as follows. In section "Method," we explained the methods used for EEG classification, including database and segmentation, data preprocessing based on WPD, and the proposed deep residual neural network. In section "Results," the numerical evaluation and experimental results of the EEG classification are shown. Finally, we give the discussion and conclusion in section "Discussion."

METHOD

Methodology Overview

The proposed EEG classification method is based on the WPD and the proposed deep residual convolutional networks. The original EEG signals were shared by the BCI Competition database (Blankertz et al., 2006). Firstly, EEG waveforms are segmented into sub-signals. Then the EEG signal features are obtained through the WPD algorithm, and some selected wavelet coefficients are retained and reconstructed into EEG signals in their respective frequency bands. Subsequently, the reconstructed EEG signals were utilized as input of the proposed deep residual convolutional networks to classify EEG signals. Finally, EEG types of motor imagination are classified by the DRes-CNN classifier intelligently.

Database and Segmentation

The international organizations have held several BCI competitions since 2001. The International BCI Competition provides a reliable data source and a unified test standard for researchers in the field of motor imaging EEG signal analysis. The experimental data in this article comes from the databases in BCI Competition 2005 (dataset IVa) and BCI Competition 2003 (dataset III).

These two databases contain data sets recorded by five subjects (aa, al, av, aw, and ay). All five subjects performed the BCI experiment, which included three exercises of motor imagination for the left hand, right hand, and right foot. In this experiment, only the right hand (R) and right foot (F) two types of motor imagination are used for data analysis, and they are named Class-1 and Class-2. Each EEG signal has 118 channels. These motor imaging tasks are classified by using the EEG signals recorded on the C3, Cz, and C4 channels. At the beginning of the experiment, a prompt appeared in the center of the screen to inform the subject of the motion imaging task to be performed. Each test takes 7 s. During the first 2 s, the subject remained sitting still. At t = 2 s, an auditory stimulus will appear, prompting the start of the experiment. At t = 3 s, an arrow will appear on the screen to indicate which imaginary exercise the subject is performing. At the same time, the subject began to perform an imaginary movement in the same direction as the arrow prompts. The subject's imagination time is 3.5 s. After the motion imaging, the subjects had a short rest period, which ranged from 1.75 to 2.25 s. At t = 7 s, the arrow disappears and the subject ends the imaginary action. The sampling frequency of EEG is 250 Hz. The EEG waveform is divided into time samples of 3.5 s. There are 140 experimental samples for each type of EEG signal.

Data Preprocessing *via* Wavelet Packet Decomposition

Electroencephalogram signals have time-varying and non-stationary characteristics. Time domain analysis mainly considers the geometric characteristics of signal variance and mean value, and frequency domain analysis mainly considers the characteristics of signal coherence and frequency band power. EEG signals are constantly changing with time. Neither time domain analysis nor frequency domain analysis alone can accurately reflect its characteristics. Time-domain joint analysis is more suitable for reflecting the transient characteristics of non-stationary signals.

Wavelet packet decomposition (Walczak and Massart, 1997; Manthalkar et al., 2003) is an improved method based on wavelet decomposition. This method makes up for the low resolution of the high-frequency part of WPD. It can analyze the signal more accurately. For different signals, WPD can automatically select the appropriate frequency band to match the frequency spectrum of the signal, thereby improving the time-frequency resolution. WPD has a good performance in signal local analysis. It can effectively remove the redundant information and retain the feature information that is beneficial to classification to best express the EEG signal feature information.

In multi-resolution analysis, WPD is regarded as a process of stepwise orthogonal decomposition of a function space. Multi-resolution analysis decomposes the $L^2(R)$ of the Hilbert space into the orthogonal sum of all wavelet subspaces $\boldsymbol{W_l}$ according to different scale factors $\boldsymbol{l}$. A new subspace $\boldsymbol{U_l^m}$ is defined to represent the wavelet subspace $\boldsymbol{W_l}$ and the scale space $\boldsymbol{V_l}$.

$$\begin{cases} \boldsymbol{U_l^0 = V_l, l \in Z} \\ \boldsymbol{U_1^1 = W_l, l \in Z} \end{cases} \tag{1}$$

The orthogonal decomposition of Hilbert space $\mathbf{V_1} \oplus \mathbf{W_1}$ can be expressed as:

$$\boldsymbol{U_l + 1^0 = U_l^0 \oplus U_l\,,\ l \in Z}\,. \tag{2}$$

Define the subspace $\boldsymbol{U_l^m}$ as a closure space of the function $\boldsymbol{u_m(t)}$, so that $\boldsymbol{u_m(t)}$ satisfies:

$$\begin{cases} \boldsymbol{u_{2m}(t) = \sqrt{2}\sum_{k\in Z} h(k)\, u_{2m}(2t-k)} \\ \boldsymbol{u_{2m+1}(t) = \sqrt{2}\sum_{k\in Z} g(k)\, u_{2m}(2t-k)} \end{cases}, \tag{3}$$

where $\boldsymbol{g(k) = (-1)^k h(1\text{-}k)}$, $\boldsymbol{g(k)}$, and $\boldsymbol{h(k)}$ are the coefficients of the high-pass filter and the low-pass filter, which are orthogonal to each other.

When m = 0, from Equation 3, we can get:

$$\begin{cases} \boldsymbol{u_0(t) = \sum_{k\in Z} h(k)\, u_0(2t-k)} \\ \boldsymbol{u_1(t) = \sum_{k\in Z} g(k)\, u_0(2t-k)} \end{cases}. \tag{4}$$

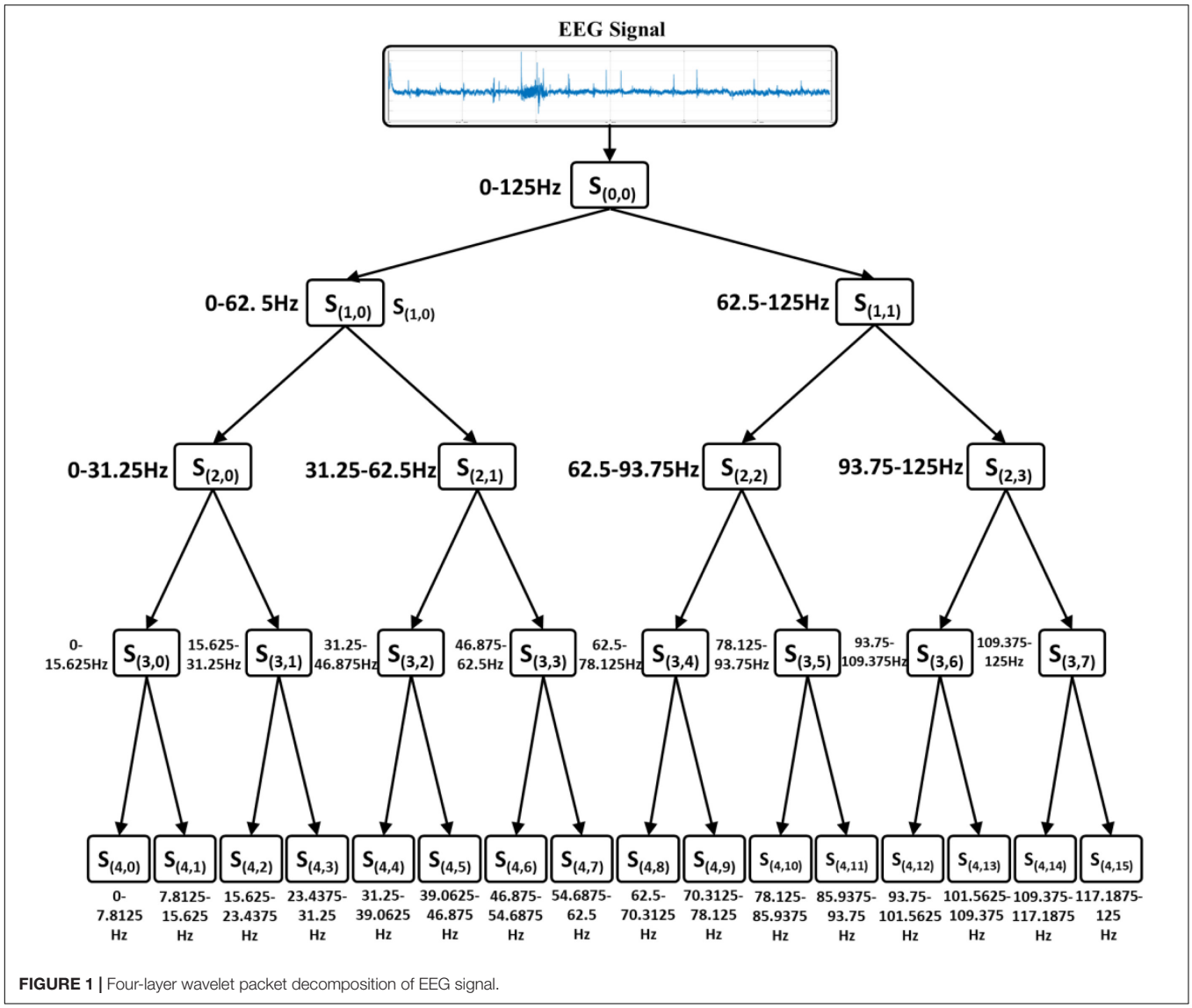


FIGURE 1 | Four-layer wavelet packet decomposition of EEG signal.

In the process of multi-resolution analysis, the wavelet basis function $\varphi(t)$ and scale function $\psi(t)$ satisfy:

$$\begin{cases} \varphi(t) = \sum_{k\in Z} \boldsymbol{g}(\boldsymbol{k})\, \psi(\boldsymbol{2t}-\boldsymbol{k}) \\ \psi(t) = \sum_{k\in Z} \boldsymbol{h}(\boldsymbol{k})\, \psi(\boldsymbol{2t}-\boldsymbol{k}) \end{cases}. \tag{5}$$

From the Equations 4, and 5, we can know that $\varphi(t) = u_1(t)$ and $\varphi(t) = u_0(t)$. Therefore $\{\boldsymbol{u_m(t)}\}_{m\in Z}$ is an orthogonal wavelet packet. The calculation formula of the WPD coefficient is shown in the Equation 6:

$$\begin{cases} d_i^{l+1,2m} = \sum_{k\in Z} \boldsymbol{h}(\boldsymbol{k}\text{-}\boldsymbol{2i})\, d_k^{l,m} \\ d_i^{l+1,2m+1} = \sum_{k\in Z} \boldsymbol{g}(\boldsymbol{k}\text{-}\boldsymbol{2i})\, d_k^{l,m} \end{cases}. \tag{6}$$

The WPD has good time-frequency resolution in both high-frequency and low-frequency parts. This method saves all the energy of the signal, so it is very suitable for the analysis and processing of EEG signals. Since the information of the EEG signal reflected by the wavelet packet coefficients on each decomposition scale is different, it can be considered to extract features from part of the wavelet packet coefficients. In the process of WPD, the decomposition scale and basis function have a great influence on the decomposition effect. The higher the scale of WPD, the better its local characteristics. But the dimensionality of the feature is also larger, which will prolong the training time of the model. Studies have confirmed that the ERD/ERS phenomenon that characterizes motor imagination actions is mainly reflected in the 8–30 Hz of the EEG signal (Pfurtscheller et al., 1997). Therefore, the decomposition scale should ensure that the frequency band corresponding to the wavelet packet coefficient is within the frequency range. Therefore, the decomposition level of WPD is determined to be four. The four-layer WPD of the EEG signal is shown in **Figure 1**. The signal is divided into 16 frequency bands, and the frequency range corresponding to each node in each layer is shown in **Figure 1**.

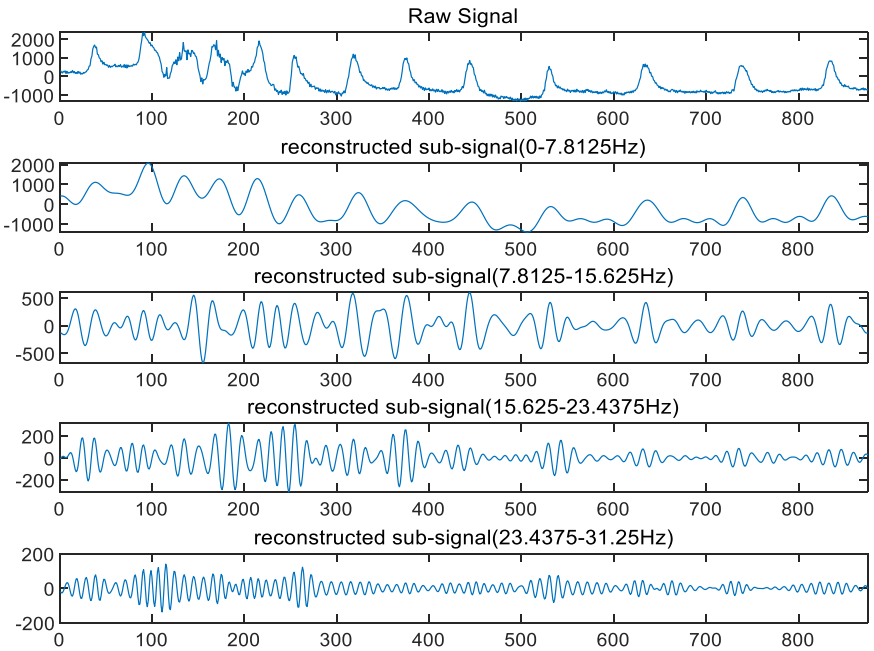


FIGURE 2 | Examples of short EEG recordings and their reconstructed sub recordings.

According to the characteristics of the original EEG signal and the wavelet basis functions, the selected wavelet basis is Db4 (Guan et al., 2015). The original EEG is decomposed by wavelet packet to obtain wavelet coefficients on various scales. The wavelet coefficients of the 0th, 1st, 2nd, and 3rd nodes in the fourth-level decomposition are retained and reconstructed into EEG signals in their respective frequency bands. The reconstructed EEG signals filter out high-frequency noise in the original signal and signals in other frequency bands that are not related to motor imagination. The reconstructed EEG signals of S_4^0, S_4^1, S_4^2, and S_4^3 are used as the input of the EEG classifier. The examples of short EEG recordings and their reconstructed sub recordings are shown in **Figure 2**.

The Proposed Deep Residual Convolutional Networks

Convolutional neural network (CNN) is a deep feedforward neural network (Liu, 2018) inspired by the concept of "receptive field." With the increase of the number of layers and neurons in the deep neural network, the non-linear fitting ability will increase. CNN is widely used in engineering fault diagnosis (Cao et al., 2019), medical signal recognition (Chandra et al., 2019; Huang et al., 2019), image recognition (Qu et al., 2016) and other fields. However, simply stacking the number of network layers will cause the problem of vanishing gradients. The network can be converged by normalizing initialization and introducing a median normalization layer. But for a deeper network, the accuracy of the model will decrease as the depth increases when the network model accuracy reaches saturation. This is the degradation problem of neural networks (He et al., 2016; Yu et al., 2016). The neural network learns an implicit abstract mapping relationship by adjusting its parameters. However, this implicit mapping relationship is difficult to optimize in a deeper network. The purpose of the deep residual convolutional neural network (DRes-CNN) method is to solve the degradation problem of traditional neural networks. The learning process of the DRes-CNN is using multiple consecutively stacked non-linear computing layers to fit the residual F(x) between the input data and the mapped output data. The residual F(x) is calculated as follows:

$$\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{x}, \tag{7}$$

where H(x) is the optimal solution, and x is the input congruent mapping.

The closer the residual F(x) is to 0, the closer the features extracted by this network are to the original input. The DRes-CNN composed of the residual block local units can solve the difficulty in convergence and adjustment. It overcomes the degradation problem of CNN as the number of network layers increases.

In this section, we propose the deep residual convolutional neural networks (DRes-CNN). As shown in **Figure 3**, the DRes-CNN is mainly composed of four residual convolution modules and a classification module. In the proposed deep residual convolutional networks, a convolutional layer with a stride of 3, a random dropout layer and a batch-normalization layer are applied firstly to compress the input EEG data and enhance the generalization of the DRes-CNN model. The down-sampling module can effectively simplifies the calculation of deep network models, reduces data redundancy, and promotes model learning (Huang et al., 2020). In the four residual convolution module, convolutional layers in series are followed by residual short circuit. Then a random dropout layer and

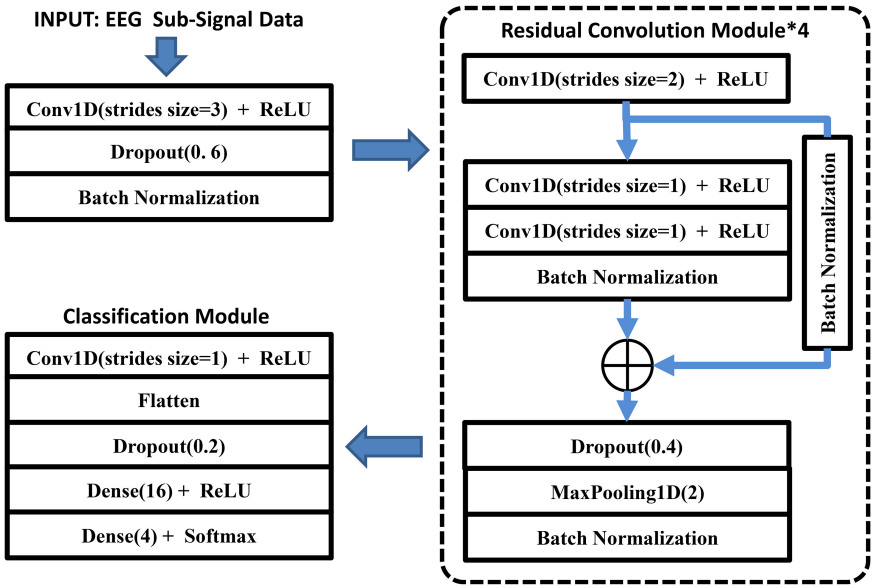


FIGURE 3 | The architecture of the proposed DRes-CNN.

the max-pooling layer are added after the convolutional layers. Finally, in the classification module, a flatten layer follows the convolution layer and a random dropout layer is applied after the flatten layer to prevent overfitting. In the proposed DRes-CNN model, the learning rate is set as 0.001 and the batch size parameter is set as 250.

RESULTS

Evaluation Metrics

The performance of the classification model is mainly measured by the accuracy. The accuracy was calculated through Equation 8.

$$Accuracy\,(\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\,, \qquad (8)$$

where TP stands for true positive, meaning the correct classification as Class-1 of EEG; TN stands for true negative, meaning correct classification as Class-2 of EEG; FP stands for false positive, meaning incorrect classification as Class-1 of EEG; FN represents false negative, meaning incorrect classification as Class-2 of EEG (Yin et al., 2016).

The Experimental Classification Results

In order to verify the effectiveness of the proposed EEG classification model, we classify the EEG signals of the right hand (Class-1) and right foot (Class-2). One hundred forty groups of eight feature inputs can be obtained for each type of EEG signal after data preprocessing based on WPD. All EEG training sample data is randomly scrambled, and the last 200 samples are selected as the test set. The classification of EEG signals is based on the classification algorithm described in section "Method."

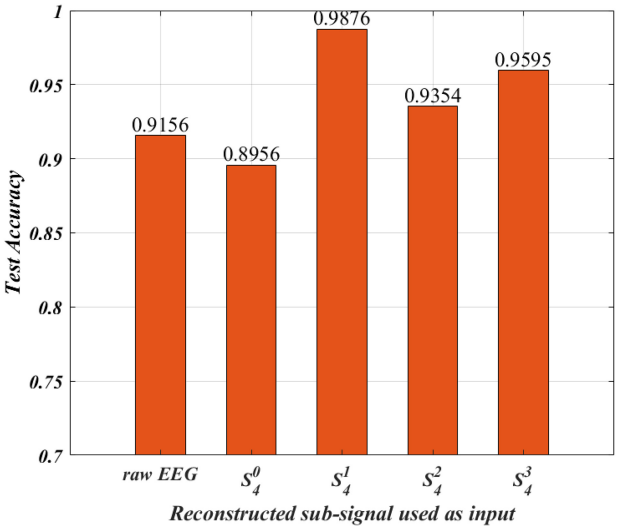


FIGURE 4 | Comparison of average test accuracy by different reconstructed sub-signal.

The original EEG waveform is divided into sub-signals. Then, the characteristics of the EEG signal are obtained through the WPD algorithm. The specific wavelet coefficients are retained and reconstructed into the EEG signals of respective frequency bands. Subsequently, the reconstructed EEG signal is used as the input of the proposed deep residual convolutional network to complete the classification of EEG signals. The experiment was run on a PC with 32 GB of memory and 16 GB of GPU memory.

Electroencephalogram signals of different frequency bands contain different characteristics and information. The wavelet

packet can decompose the information of each frequency band. It makes the characteristics of the EEG signal easier to identify in each frequency band. In this section, the performance of training the DRes-CNN with reconstructed sub-signals as model input is studied experimentally, and the experiment results are shown in **Figure 4**.

From **Figure 4**, we can find that the accuracy of the classification test using the original EEG signal as input reaches 91.56%. The reconstructed EEG data set of S_4^0 (0, 7.8125 Hz) reached an average test accuracy of 89.56%. The reconstructed EEG data set of S_4^1 (7.8125, 15.625 Hz) reached the best average test accuracy of 98.76%. The reconstructed EEG data of S_4^2 (15.625, 23.4375 Hz) set reached an average test accuracy of 93.54%. The reconstructed EEG data set of S_4^3 (23.4375, 31.25 Hz) reached an average test accuracy of 95.95%. From the experimental comparison demonstrated above, we can conclude that the proposed DRes-CNN model shows the best classification performance when the reconstructed EEG dataset of S_4^1 (7.8125, 15.625 Hz) is used as model input.

DISCUSSION

In this article, we proposed an EEG classification method using WPD and the proposed deep residual convolutional network. The goal of this method is to achieve high-precision intelligent classification of motor-imaging EEG signals. The original EEG signal is shared by the BCI Competition database. Firstly, the EEG waveform is divided into shorter sub-signals. Then, the characteristics of the EEG signal are obtained through the WPD algorithm. Some selected wavelet coefficients are retained and reconstructed into EEG signals of their respective frequency bands. Subsequently, the reconstructed EEG signal is used as the input of the proposed deep residual convolutional network. Finally, the motor imagery EEG signals are intelligently classified by the DRes-CNN classifier. We compared the classification performance of reconstructed signals in different frequency bands as input to the model. Through comparative experiments, we found that the proposed DRes-CNN model shows the best classification accuracy of 98.76% when the reconstructed EEG data set in the frequency band of (7.8125, 15.625 Hz) is used as the model input. The proposed method can be further applied to the BCI system of motor imagination control.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

J-SH, W-SL, and BY conceived and designed the classification method. Z-XW and S-FC performed the experiment. J-SH preprocessed and analyzed the data and wrote the manuscript. BY, S-FC, and W-FS reviewed and edited the manuscript. J-SH and W-FS responded to the comments of the reviewers. All authors read and approved the manuscript.

FUNDING

This research was supported financially by the National Natural Science Foundation of China (No. 51605403), the Fundamental Research Funds for the Central Universities under Grant 20720190009, International Science and Technology Cooperation Project of Fujian Province of China under Grant 2019I0003, China Aviation Engine Corporation's 2019 Industry-University-Research Cooperation Project Funding Project (No. HFZL2019CXY02), Natural Science Foundation of Fujian Province (2020J02063), and Xiamen Science and Technology Bureau Foundation of Science and Technology Project for Medical and Healthy (3502Z20209005).

REFERENCES

Anumanchipalli, G., Chartier, J., and Chang, E. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature* 568, 493–498. doi: 10.1038/s41586-019-1119-1

Bai, X. (2010). Real-time control of electric wheelchairs by brain waves. *Robot Technol. Appl.* 2, 10–12.

Benjamin, B., Guido, D., Matthias, K., Schroder, M., Williamson, J., Murray-Smit, R., et al. (2003). The Berlin brain-computer interface presents the novel mental typewriter Hex-O-Spell. *Comput. Intell. Neurosci.* 14, 332–336.

Blankertz, B., Muller, K. R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlögl, A., et al. (2006). The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159. doi: 10.1109/TNSRE.2006.875642

Cao, X. C., Chen, B. Q., Yao, B., and He, W.-P. (2019). Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification. *Comput. Ind.* 106, 71–84. doi: 10.1016/j.compind.2018.12.018

Chandra, B. S., Sastry, C. S., and Jana, S. (2019). Robust heartbeat detection from multimodal data via CNN-Based generalizable information fusion. *IEEE Trans. Biomed. Eng.* 66, 710–717. doi: 10.1109/TBME.2018.2854899

Cilliers, P. J., and Van Der Kouwe, A. J. W. (1993). "A VEP-based computer interface for C2-ouadriplegics," in *Proceedings of the 15th Annual International Conference of the IEEE*, (San Diego, CA: IEEE), 1263–1263. doi: 10.1109/IEMBS.1993.979126

Gong, P. (2014). *Research on Feature Extraction of Motor Imagination EEG Signal*. Chongqing: Chongqing University.

Guan, W., Qingzhi, Y., Yang, G., et al. (2015). Detection and location of transient power quality disturbances in distribution network based on db4 wavelet. *Power Syst. Prot. Control* 43, 102–106.

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*, Las Vegas, NV, 770–778. doi: 10.1109/CVPR.2016.90

Huang, J., Chen, B., Yao, B., and He, W. (2019). ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access* 7, 92871–92880. doi: 10.1109/ACCESS.2019.2928017

Huang, J., Chen, B., Zeng, N., Cao, X.-C., and Li, Y. (2020). Accurate classification of ECG arrhythmia using MOWPT enhanced fast compression deep learning networks. *J. Ambient Intell. Humaniz. Comput.* 5, 1–18. doi: 10.1007/s12652-020-02110-y

Kemal, P., and Salih, G. (2007). Classification of epileptiform eeg using a hybrid system based on decision tree classifier and fast fourier

transform. *Appl. Math. Comput.* 187, 1017–1026. doi: 10.1016/j.amc.2006.09.022

Li, Y., Long, J., Yu, T., Yu, Z., Wang, C., Zhang, H., et al. (2010). An EEG-based BCI system for 2-D cursor control by combining Mu/Beta rhythm and P300 potential. *IEEE Trans. Biomed. Eng.* 57, 495–2505. doi: 10.1109/TBME.2010.2055564

Liu, C. (2018). *Research and Design of Handwritten Digit Recognition Based on Convolutional Neural Network*. Chengdu: Chengdu University of Technology.

Manthalkar, R., Biswas, P. K., and Chatterji, B. N. (2003). Rotation and scale invariant texture features using discrete wavelet packet transform. *Patt. Recognit. Lett.* 24, 2455–2462. doi: 10.1016/S0167-8655(03)00090-4

McFarland, D. J., Sarnacki, W. A., and Wolpaw, J. R. (2010). Electroencephalographic (EEG) control ofthree-dimensional movement. *J. Neural Eng.* 7:036007. doi: 10.1088/1741-2560/7/3/036007

Mervyn, V., Li, X., Shen, K., and Wilder-Smith, E. P. (2009). Can SVM be used for automatic EEG detection of drowsiness during car driving. *Saf. Sci.* 47, 115–124. doi: 10.1016/j.ssci.2008.01.007

Nguyen, T., Nguyen, T., Truong, K., and Toi, V. (2013). *A Mean Threshold Algorithm for Human Eye Blinking Detection Using EEG*. Berlin: Springer, 275–279. doi: 10.1007/978-3-642-32183-2_69

Pfurtscheller, G., Flotzinger, D., and Kalcher, J. (1993). Brain-computer interface-a new communication device for handicapped persons. *J. Microcomput. Appl.* 16, 293–299. doi: 10.1006/jmca.1993.1030

Pfurtscheller, G., Muller-Putz, G. R., Schlogl, A., et al. (2006). 15 years of BCI research at graz university of technology : current projects. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 205–210. doi: 10.1109/TNSRE.2006.875528

Pfurtscheller, G., Neuper, C., Flotzinger, D., and Pregenzer, M. (1997). EEG-based discrimination between imagination of right and left hand movement. *Electroencephalogr. Clin. Neurophysiol.* 103, 642–651. doi: 10.1016/S0013-4694(97)00080-1

Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W., Ramoser, H., Schlögl, A., et al. (2000). Current trends in Graz brain–computer interface (BCI) research. *IEEE Trans. Rehabil. Eng.* 8, 216–219. doi: 10.1109/86.847821

Qu, J., Xian, S., and Xin, G. (2016). Remote sensing image target recognition based on CNN. *Foreign Electronic Meas. Technol.* 8, 45–50.

Satapathy, S., Jagadev, A., and Dehuri, S. (2017). Weighted majority voting based ensemble of classifiers using different machine learning techniques for classification of eeg signal to detect epileptic seizure. *Informatica* 41, 99–110.

Walczak, B., and Massart, D. L. (1997). Noise suppression and signal compression using the wavelet packet transform. *Chemometr. Intell. Lab. Syst.* 36, 81–94. doi: 10.1016/S0169-7439(96)00077-9

Wang, T., Guan, S., Man, K., and Ting, T. (2014). EEG eye state identification using incremental attribute learning with time-series classification. *Math. Probl. Eng.* 2014:365101. doi: 10.1155/2014/365101

Wang, Y., Hong, B., Gao, X., and Gao, S. (2007). Implementation of a Brain-computer interface based on three states of motor imagery. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2007, 5059–5062. doi: 10.1109/IEMBS.2007.4353477

Xu, B., Aiguo, S., and Renhuan, Y. (2011). Online brain-computer interface experiment based on motor imagination EEG. *J. Huazhong Univ. Sci. Technol.* 39, 60–64.

Yin, W., Yang, X., Zhang, L., and Oki, E. (2016). ECG monitoring system integrated with IR-UWB radar based on CNN. *IEEE Access* 4, 6344–6351. doi: 10.1109/ACCESS.2016.2608777

Yu, L., Chen, H., Dou, Q., Qin, J., and Heng, P.-A. (2016). Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans. Med. Imaging* 36, 994–1004. doi: 10.1109/TMI.2016.2642839

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

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

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

frontiers in Neuroscience

TECHNOLOGY AND CODE
published: 23 November 2021
doi: 10.3389/fnins.2021.764796

Vector Auto-Regressive Deep Neural Network: A Data-Driven Deep Learning-Based Directed Functional Connectivity Estimation Toolbox

Takuto Okuno and*
Alexander Woodward for the Alzheimer's Disease Neuroimaging Initiative†

Connectome Analysis Unit, RIKEN Center for Brain Science, Wako, Japan

OPEN ACCESS

Edited by:
Mohammad Khosravi, Persian Gulf University, Iran

Reviewed by:
Mark William Woolrich, University of Oxford, United Kingdom
Xiaotong Wen, Renmin University of China, China

***Correspondence:**
Takuto Okuno
takuto.okuno@riken.jp

†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *26 August 2021*
Accepted: *19 October 2021*
Published: *23 November 2021*

Citation:
Okuno T and Woodward A (2021) Vector Auto-Regressive Deep Neural Network: A Data-Driven Deep Learning-Based Directed Functional Connectivity Estimation Toolbox. Front. Neurosci. 15:764796. doi: 10.3389/fnins.2021.764796

An important goal in neuroscience is to elucidate the causal relationships between the brain's different regions. This can help reveal the brain's functional circuitry and diagnose lesions. Currently there are a lack of approaches to functional connectome estimation that leverage the state-of-the-art in deep learning architectures and training methodologies. Therefore, we propose a new framework based on a vector auto-regressive deep neural network (VARDNN) architecture. Our approach consists of a set of nodes, each with a deep neural network structure. These nodes can be mapped to any spatial sub-division based on the data to be analyzed, such as anatomical brain regions from which representative neural signals can be obtained. VARDNN learns to reproduce experimental time series data using modern deep learning training techniques. Based on this, we developed two novel directed functional connectivity (dFC) measures, namely VARDNN-DI and VARDNN-GC. We evaluated our measures against a number of existing functional connectome estimation measures, such as partial correlation and multivariate Granger causality combined with large dimensionality counter-measure techniques. Our measures outperformed them across various types of ground truth data, especially as the number of nodes increased. We applied VARDNN to fMRI data to compare the dFC between 41 healthy control vs. 32 Alzheimer's disease subjects. Our VARDNN-DI measure detected lesioned regions consistent with previous studies and separated the two groups well in a subject-wise evaluation framework. Summarily, the VARDNN framework has powerful capabilities for whole brain dFC estimation. We have implemented VARDNN as an open-source toolbox that can be freely downloaded for researchers who wish to carry out functional connectome analysis on their own data.

Keywords: vector auto-regressive deep neural network, directed functional connectivity, fMRI, Granger causality, Alzheimer's disease

INTRODUCTION

There is a rich history of structural and functional connectome analysis of neuroimaging data of the whole brain, such as those acquired from MRI (magnetic resonance imaging). One important goal is to study the causal relationship between activity in different brain regions and the brain's functional circuitry. The estimation of the functional connectome is an important step for understanding brain function (Gao et al., 2011; Krueger et al., 2011; Hahn et al., 2019) and for the diagnosis of brain lesions (Dauwels et al., 2010). In the literature, there is a notable lack of data-driven approaches to functional connectome estimation that leverage the advantages of the current state-of-the-art deep neural network architectures and training methodologies. Such approaches should have an advantage in terms of representational power and trainability. Some recent examples of neural network-based measures of GC have been proposed, such as neural network GC (NN-GC) (Montalto et al., 2015), RNN-GC (Wang et al., 2018), Echo State Network GC (ES-GC) (Duggento et al., 2019), and DNN-GC (Chivukula et al., 2018). Accordingly, we propose a new approach for whole-brain analytics based on a *vector auto-regressive deep neural network* (VARDNN) architecture that can deal with a large number of time series. The model comprises a set of nodes with a deep neural network (DNN) structure. Each node maps with arbitrary divisions of space, such as an anatomical brain region from where a representative neural signal (time-series) can be estimated. At each time step, a node outputs a signal by considering exogenous signals into the system, signals from all other nodes, and feedback from the node back to itself. The model can be trained on time series data of brain activities and recordings of arbitrary exogenous inputs. Functional connectome measures can be derived from the model after the training process. A wide variety of data can be used for training, such as fMRI BOLD or calcium imaging data; our approach is fully data-driven. We seek a non-parametric model that not only fits the data but also describes important dynamical properties and can complement hypothesis-driven approaches (Valdes-Sosa et al., 2011). We defined two types of *directed functional connectivity* (dFC) measures based on the VARDNN, namely VARDNN-DI (Directional Influence, our new measure) and VARDNN-GC (based on Granger causality).

Many types of functional connectome analysis algorithms are currently used. *Functional connectivity* (FC), which is simply calculated by the correlation coefficient of *pairwise* time-series, is the most basic measure to analyze brain region relationships. This term was originally defined as the temporal correlation of a neurophysiological index measured in different brain areas (Gerstein et al., 1989; Friston et al., 1993). *Partial Correlation* (PC) is a *conditional* version of correlation. This is calculated by the correlation of regression residuals with other time-series. When two factors show a strong correlation, it may be necessary to consider the presence of a third factor. PC can remove this kind of third factor effect and show a normalized correlation. FC based on both correlation and partial correlation only consider the *zero-lag* relationship between signals. On the other hand, *Granger causality* (GC) is based on a vector auto-regression (VAR) scheme and is categorized as a *predictive* type measure. The original formulation of GC was for the *pairwise* case but has since been extended to the *conditional* and *multivariate* cases (Granger, 1969; Geweke, 1982; Goebel et al., 2003). Furthermore, an information-theoretic approach to understanding causal relationships, called *transfer entropy* (TE), has been developed (Schreiber, 2000). This has been shown to be equivalent to GC when the time series dynamics under study are normally distributed (Barnett et al., 2009). GC and TE are directional, thus these measures are categorized as *directed functional connectivity* (dFC) (Friston et al., 2013), while FC and PC are categorized as *undirected functional connectivity*.

Some measures depend on a linear regression scheme, thus their coefficients and residuals are calculated by ordinary least squares (OLS). But OLS has several drawbacks: if the predictor variables (P) are larger than the observed data (N), the solution faces overfitting ($P >> N$ problem). Furthermore, if the observed data are strongly correlated with each other, the Gram matrix cannot be inverted and OLS cannot calculate the correct solution (multicollinearity problem). In the case of analyzing functional MRI data, sometimes several regions have strong correlations, or the observed data may be smaller than the region number. Here, functional connectome analysis algorithms face several OLS problems. To avoid these situations, countermeasure techniques can be applied. For example, PC has been combined with a L1-norm regularized regression scheme (SPC-EN) (Ryali et al., 2012), and multivariate GC has been combined with Principal Component Analysis (PCA-cGCM) (Zhou et al., 2009) to reduce the data dimension and transform the data into an orthogonal vector space.

Accurate estimation of the brain's functional connectome is an important topic for neuroscience. One challenge is that the estimated functional connectome values are not consistent across analysis algorithms, which can cause confusion when interpreting the results of a particular measure. Dauwels et al. (2010) attempted to systematically investigate EEG (electroencephalogram) synchrony with a special focus on the early diagnosis of Alzheimer's disease (AD). They evaluated 29 types of functional measures, including the correlation coefficient and related measures, phase synchrony indices, GC, and stochastic event synchrony (SES) (Dauwels et al., 2008). Results showed that only two measures, GC and SES, were able to convincingly distinguish MCI (mild cognitive impairment) patients from age-matched controls. Another investigation was performed by Smith et al. (2011). They evaluated 20 functional methods, including the correlation coefficient and related measures, mutual information (Shannon, 1948), GC, the LiNGAM algorithm (Shimizu et al., 2006), and several Bayes Net modeling algorithms. Twenty-eight types of synthetic fMRI BOLD (blood oxygen level dependent) signals were generated by the DCM model (Friston et al., 2003), and all methods were evaluated for their estimation of the ground truth connectivity. In the results, the PC regularized inverse covariance estimation, and several Bayes Net methods showed the highest sensitivity. Prando et al. (2020) also evaluated the performance of functional connectome estimation using a number of methods. They compared sparse vs. spectral

DCM, multivariate GC, LiNGAM, and several Bayes Net modeling algorithms. In total, 9 methods were evaluated using synthetic fMRI BOLD signals generated by the DCM model. The results showed that the sparse DCM method gave the highest accuracy and multivariate GC was competitive. Overall, Granger causality-based measures appear to give consistent performance for functional connectome estimation from both empirical and synthetic signals. A purely data-driven approach makes less assumptions about the underlying system generating the dynamics; hence, it can be applied more broadly to different types of phenomena and is less computationally expensive.

To evaluate our VARDNN approach we tested the performance of its dFC measures against 14 other analysis algorithms, including *zero-lag* and *predictive* types, and *pairwise* and *multivariate* strategies. Included in these algorithms, countermeasure techniques, such as PCA, Elastic Net (Zou and Hastie, 2005) and Partial Least Squares (PLS) (Wold et al., 2001) were combined with PC and multivariate GC. All measures were evaluated on biologically plausible synthetic signals generated by DCM and the reduced Wong-Wang neural mass model (Wong and Wang, 2006; Deco et al., 2013). In particular, we found that both measures, especially VARDNN-DI, were able to consistently provide competitive and top-ranking performance, especially when the size of the number of nodes increased. Furthermore, we evaluated VARDNN performance by analyzing empirical fMRI BOLD data. Healthy control and Alzheimer's disease fMRI BOLD data were acquired from the *Alzheimer's Disease Neuroimaging Initiative* (ADNI2) database, and functional connectomes were calculated by each of the analytical methods studied. Functional connectome values were then used to compare group separation ability of Alzheimer's disease within a subject-wise evaluation framework. Functional connectomes estimated by the VARDNN measure was able to capture the causal relationship between brain regions and showed a significant difference between healthy control and Alzheimer's disease data.

MATERIALS AND METHODS

Vector Auto-Regressive Deep Neural Network Architecture

Figure 1 shows a schematic diagram of the VARDNN framework. The architecture consists of a set of nodes that can be assigned to different regions of the brain from which experimental signals are recorded. The node state vector is $\boldsymbol{S}(t) = (S_1(t), \cdots, S_n(t))$, where $S_i(t) \in \boldsymbol{R}$, $i = 1,..,n$, represents the ith node's state at discrete time $t = 1,2,\ldots$. All nodes are part of a fully connected network structure, receiving input from every other node. Additionally, each node receives exogenous signals from outside of the brain, $I_q(t) \in \boldsymbol{R}$, $q = 1,\ldots,m$, expressed as an exogenous signal vector $\boldsymbol{I}(t) = (I_1(t), \cdots, I_m(t))$. These signals can arbitrarily represent any external data derived

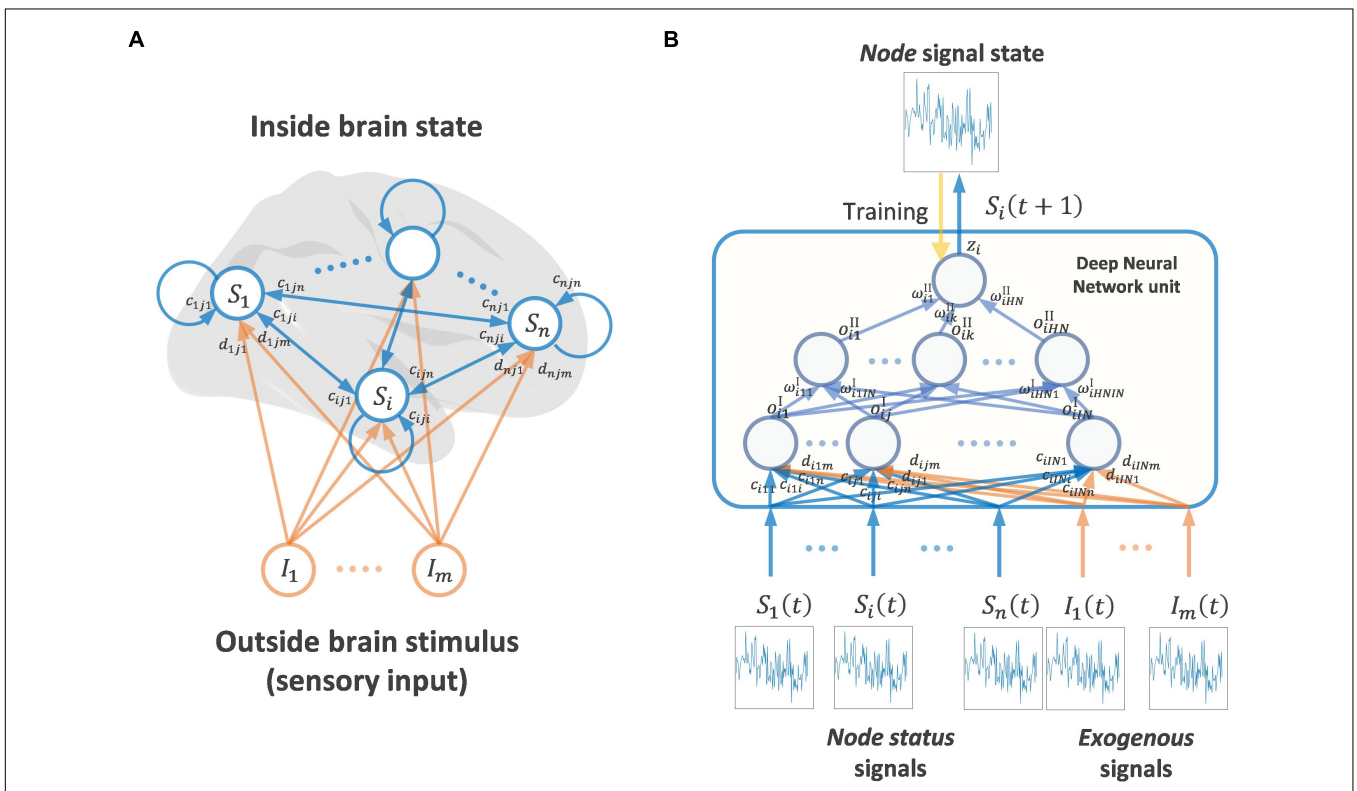


FIGURE 1 | (A) Whole brain node state network and stimulus coming from outside brain (mainly sensory input) **(B)** Showing signaling framework of one node state. A Deep Neural Network node receives the signal state of other brain nodes and exogenous signals (outside brain stimulus). A DNN unit calculates one output signal, in this case the next time step of the signal state of a node. The DNN unit is trained by teacher signals and the input of other node states and exogenous signals.

from an experimental paradigm that would be received as sensory information by the subject. All nodes contain a DNN unit with trainable weights across two hidden layers. For a node i, $c_{i,j,p}, d_{i,j,q} \in \boldsymbol{R}$ are the connection weights between its first hidden layer neurons $x_{i,j}$, $j = 1,\ldots,IN$, and the input from every other deep neural network unit node $p = 1,\ldots,n$ (including its self-connection), or from the exogenous signal inputs ($q = 1,\ldots,m$), respectively. At each time step, signals pass between nodes and are processed by the DNN units. Within a DNN, the output of the first hidden layer for node i is expressed as

$$\boldsymbol{O}_i^{\mathrm{I}} = g(\boldsymbol{C}_i \boldsymbol{S}(t)^{\mathrm{T}} + \boldsymbol{D}_i \boldsymbol{I}(t)^{\mathrm{T}} + \boldsymbol{B}_i{}^{\mathrm{T}})^{\mathrm{T}} \quad (1)$$

where

$$\boldsymbol{C}_i = \begin{pmatrix} c_{i,1,1} & \cdots & c_{i,1,n} \\ \vdots & \ddots & \vdots \\ c_{i,IN,1} & \cdots & c_{i,IN,n} \end{pmatrix}, \boldsymbol{D}_i = \begin{pmatrix} d_{i,1,1} & \cdots & d_{i,1,m} \\ \vdots & \ddots & \vdots \\ d_{i,IN,m} & \cdots & d_{i,IN,m} \end{pmatrix},$$

$$\boldsymbol{B}_i = \left(b_{i,1}, \cdots, b_{i,IN}\right)$$

where $b_{i,j} \in \boldsymbol{R}$ are the bias values on the first hidden layer neurons and $g(x) = \max(0, x)$ is the non-linear ReLU activation function (Nair and Hinton, 2010; Glorot et al., 2011).$\boldsymbol{O}_i^{\mathrm{I}}$ is passed to the second hidden layer neurons of node i, and its output is expressed as

$$\boldsymbol{O}_i^{\mathrm{II}} = g(\boldsymbol{W}_i^{\mathrm{I}} \boldsymbol{O}_i^{\mathrm{I}^{\mathrm{T}}} + \boldsymbol{\beta}_i^{\mathrm{I}^{\mathrm{T}}})^{\mathrm{T}} \quad (2)$$

where

$$\boldsymbol{W}_i^{\mathrm{I}} = \begin{pmatrix} \omega_{i,1,1}^{\mathrm{I}} & \cdots & \omega_{i,1,IN}^{\mathrm{I}} \\ \vdots & \ddots & \vdots \\ \omega_{i,HN,1}^{\mathrm{I}} & \cdots & \omega_{i,HN,IN}^{\mathrm{I}} \end{pmatrix}, \boldsymbol{\beta}_i^{\mathrm{I}} = (\beta_{i,1}^{\mathrm{I}}, \cdots, \beta_{i,HN}^{\mathrm{I}})$$

$\omega_{i,j,k}^{\mathrm{II}} \in \boldsymbol{R}$ are the weight values of the second hidden layer neurons $k = 1,\ldots,HN$, and $\beta_{i,k}^{\mathrm{II}} \in \boldsymbol{R}$ are the biases. Finally, node i's DNN second hidden layer outputs are processed by an output layer neuron:

$$z_i(t) = \boldsymbol{W}_i^{\mathrm{II}} \boldsymbol{O}_i^{\mathrm{II}^T} + \beta_i^{\mathrm{II}} \quad (3)$$

where $\boldsymbol{W}_i^{\mathrm{II}} = (\omega_{i,1}^{\mathrm{II}}, \cdots, \omega_{i,HN}^{\mathrm{II}})$ is the output neuron's weight vector, and $\beta_i^{\mathrm{II}} \in \boldsymbol{R}$ is its bias. Finally, the VARDNN's new state at $t+1$ for all nodes is taken as the value of each node's output:

$$\boldsymbol{S}(t+1) = \boldsymbol{Z}(t) = (z_1(t), \cdots, z_n(t)).$$

For our current implementation, we defined 2 hidden layers for each DNN unit in the VARDNN architecture. However, the DNN unit is not limited to 2 hidden layers, and other forms of DNNs can also be used in the framework. Our decision was based on this work's focus on analyzing whole brain fMRI BOLD type data, and 2 layers were empirically determined (see **Supplementary Figure 1** for a description of the procedure). Although neural networks with single hidden layers are known as universal approximators (Cybenko, 1989; Hornik, 1989), a single hidden layer requires an exponential increase in the number of hidden neurons proportional to the problem size (Goodfellow et al., 2016; Liang and Srikant, 2016). Theoretically and experimentally, deep hidden layers with smaller numbers of neurons have shown higher accuracy than shallow (single) hidden layer networks (Delalleauand Bengio, 2011; Goodfellow et al., 2014).

VARDNN can learn to reproduce specific time series data through training on teacher data. The teacher data can be defined as $\boldsymbol{Teach}(t) = \left(S'_1(t), \cdots, S'_n(t), I'_1(t), \cdots, I'_m(t)\right)$, where S'_i and I'_i are the experimentally obtained signals of "neural activity" and (if any) exogenous inputs, respectively, for each node. Here, t = 1,2,...,L, where L is the number of *frames* of data at a sampling period TR. The number of nodes, n, corresponds to the number of input time series. This number is arbitrary and should reflect some spatial sub-division of the brain, based on the nature of the teacher data. For example, if fMRI BOLD data is used, then one can average the voxel values within each anatomical region to generate a set of representative teacher signals.

Each node is independently trained with the aim to iteratively reduce the error between the VARDNN-generated $t+1$ state and the teacher data at time t+1, when the current state is set by the time t teacher data. With teacher data of length L frames, the training set for a node i comprises $\{\boldsymbol{\phi}_i(1), \ldots, \boldsymbol{\phi}_i(L-1)\}$, where $\boldsymbol{\phi}_i(k) = \{\boldsymbol{Teach}(k), S'_i(k+1)\}$ is an input/target pair, k = 1,2,...,L-1. Even though these pairs are derived from time series data, they can be used independently. Thus, the shuffling and minibatch techniques can be applied in each training epoch. This feature of VARDNN is clearly different from recurrent neural network (RNN) architectures as VARDNN is simply trained by input/target patterns but not sequences.

Within each training epoch, for each training pair $\boldsymbol{\phi}_i(k)$, the VARDNN node state and exogenous input are set to $\boldsymbol{Teach}(k)$, and the output scalar value $z_i(k)$ is calculated. Then, the error is evaluated as the difference $z_i(k) - S'_i(k+1)$. The training error for node i is expressed over all $L-1$ training pairs as $\boldsymbol{Err}_i = \boldsymbol{y}_i - \widetilde{\boldsymbol{z}}_i$, where $\boldsymbol{y}_i = (S'_i(2), \ldots, S'_i(L))$ and $\widetilde{\boldsymbol{z}}_i = (z_i(1), \cdots, z_i(L-1))$. This error is used to drive the optimization procedure.

After a number of epochs, the training is completed for the current node, and this process is repeated for all nodes of the VARDNN system. Currently, the VARDNN framework was implemented in MATLAB using the Deep Learning Toolbox version 12.1. Training was optimized using Adam (Kingma and Ba, 2014) with L2 regularization for the hidden layers to avoid over-fitting.

Vector Auto-Regressive Deep Neural Network-Granger Causality Measure

Granger causality is a well-established method for the causal analysis of time series data, and the multivariate version (mvGC) has successfully been used to analyze fMRI BOLD signals

(Goebel et al., 2003; Menon and Uddin, 2010; Marinazzo et al., 2011; Wen et al., 2012). The traditional approach to multivariate Granger causality uses a vector auto-regressive model involving a state vector $\boldsymbol{S}(t)$, time lag p, coefficient vector $\boldsymbol{A}_{ik}$, and residual $\varepsilon_i\ (t)$. The state of node i can be described by the following model:

$$S_i(t) \ = \ \sum_{k=1}^{p} \boldsymbol{A}_{ik} \boldsymbol{S}\left(t-k\right)^{\mathrm{T}} + \varepsilon_i\ (t). \tag{4}$$

Removing the past dependency of node i on j, the model can be written as

$$S_{i\backslash \mathrm{j}}(t) \ = \ \sum_{k=1}^{p} \boldsymbol{A}_{ik\backslash \mathrm{j}} \boldsymbol{S}_{i\backslash \mathrm{j}} \left(t-k\right)^{\mathrm{T}} + \varepsilon_{i\backslash \mathrm{j}}(t). \tag{5}$$

Then the mvGC causal index is expressed as

$$\mathcal{F}_{j\rightarrow i} \ = \ \log\left(var\left(\boldsymbol{\varepsilon}_{i\backslash \mathrm{j}}\right)/var\left(\boldsymbol{\varepsilon}_i\right)\right) \tag{6}$$

where $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_{i\backslash \mathrm{j}}$ are time series of residuals, and $var()$ denotes the statistical variance. For mvGC, the coefficients and residuals of the vector auto-regressive model are calculated by OLS. However, as mentioned in section "Introduction," neural network-based measures of GC have recently been proposed. VARDNN can also calculate this type of Granger causality. We simply use $\boldsymbol{Err}_i = \boldsymbol{y}_i - \widetilde{\boldsymbol{z}}_i$ of a single node i as the residuals $\boldsymbol{\varepsilon}_i$ and the error when removing the signal generated from node j from the optimization, $\boldsymbol{Err}_{i\backslash \mathrm{j}} = \boldsymbol{y}_i - \widetilde{\boldsymbol{z}}_{i\backslash \mathrm{j}}$ as the residuals $\boldsymbol{\varepsilon}_{i\backslash \mathrm{j}}$. Then, the VARDNN-GC measure is defined as

$$\mathcal{F}_{j\rightarrow i}^{VARDNN-GC} \ = \ \log\left(var\left(\boldsymbol{Err}_{i\backslash \mathrm{j}}\right)/var\left(\boldsymbol{Err}_i\right)\right) \tag{7}$$

In the context of neuroscience, the process of removing the signal from another node can be thought of as performing a simulated lesion (Honey et al., 2007; Alstott et al., 2009). This virtual impairment can allow us to understand the effect of possible lesions in terms of their (Granger causal) connectivity relationships. Therefore, VARDNN can also function as a lesion model.

Vector Auto-Regressive Deep Neural Network-Directional Influence Measure

Vector auto-regression is a well-known statistical approach to model the relationship between a scalar response and one or more explanatory variables. From this perspective, we can describe the state of a node i as

$$\begin{aligned} S_i(t) &= \alpha_{i,0} \ + \ \boldsymbol{\alpha}_i \boldsymbol{S}(t-1)^{\mathrm{T}} \ + \ \varepsilon_i(t) \\ &= \alpha_{i,0} \ + \ \sum_{k=1}^{n} \alpha_{i,k} S_k(t-1) \ + \ \varepsilon_i(t) \end{aligned} \tag{8}$$

with state vector $\boldsymbol{S}(t)$, linear relationship vector $\boldsymbol{\alpha}_i = \left(\alpha_{i,1}, \cdots, \alpha_{i,n}\right)$ and a residual term $\varepsilon_i\ (t)$. Now, if we set all inputs into the node as 1, in other words $\boldsymbol{S}(t-1) = (1, \cdots, 1)$ we get

$$S_i^{\mathrm{one}} \ = \ \sum_{k=0}^{n} \alpha_{i,k} \ + \ \varepsilon_i(t) \tag{9}$$

If input to node j is set to 0, $S_{i\backslash \mathrm{j}}^{\mathrm{one}}$ is expressed as

$$S_{i\backslash \mathrm{j}}^{\mathrm{one}} \ = \ \sum_{k\neq \mathrm{j}} \alpha_{i,k} \ + \ \varepsilon_i(t) \tag{10}$$

Then, the magnitude of the linear relationship, $\alpha_{i,j}$, or *directional influence* (DI), from node j to i is expressed as

$$\alpha_{i,j} \ = \ S_i^{\mathrm{one}} - S_{i\backslash \mathrm{j}}^{\mathrm{one}} \tag{11}$$

The VARDNN formula is complex, but non-linearly extended from this VAR formulation. After training the VARDNN, the deep neural network weights of each node encode its non-linear relationships with all other nodes. In a similar manner to the linear case, we defined a new measure of *directional influence*, VARDNN-DI, that extracts the non-linear magnitude of causal relationships from the network weights. Here, the DI from node j to node i is defined as

$$\mathcal{F}_{j\rightarrow i}^{VARDNN-DI} \ = \ \left| z_i^{\mathrm{one}} - z_{i\backslash \mathrm{j}}^{\mathrm{one}} \right|. \tag{12}$$

In this case, z_i^{one} is calculated by setting all inputs into the node to 1, that is $\boldsymbol{S}(t) = (1, \cdots, 1)$ and $\boldsymbol{I}(t) = (1, \cdots, 1)$, and passing these values through its DNN. $z_{i\backslash \mathrm{j}}^{\mathrm{one}}$ is also calculated in the same way but with the connection weight from node j set zeros in the first hidden layer neurons of the DNN of node i. As a result, the learned non-linear relationship from node j to node i can be extracted.

Pre-processing of Vector Auto-Regressive Deep Neural Network Teacher Data

Although it is not a restriction, the VARDNN was designed to preferably deal with data in the range [0, 1]. In practice it may occur that the teacher signals are not normalized to this range. For example, resting state fMRI (rsfMRI) BOLD data has a normal distribution centered around zero (Chen et al., 2003). Thus, we need to transform experimentally derived time series data into the [0, 1] range. To do this, the mean and standard deviation of the data are first estimated:

$$\mu^* \ = \ \frac{1}{nL} \sum_{i=1}^{n} \sum_{t=1}^{L} S_i(t), \tag{13}$$

$$\sigma^* \ = \ \sqrt{\frac{1}{nL} \sum_{i=1}^{n} \sum_{t=1}^{L} \left(S_i(t) - \mu^*\right)^2}. \tag{14}$$

The values of a particular time series, mapping to a node i are normalized by

$$\theta_i(t) = \left(S_i(t) - \mu^*\right)/\sigma^*. \quad (15)$$

Then, a sigmoid function is used to transform the signal into the [0, 1] range:

$$S_i'(t) = 1/\left(1 + e^{-\alpha\theta_i(t)}\right), \quad (16)$$

where the coefficient $\alpha = 1$ was used for our current experiments. This transformation was chosen as it deals with potential signal outliers (that, for example, can appear in fMRI BOLD data) better than a simple linear transformation (see **Supplementary Figure 2** for an evaluation of the effect of sigmoid transformation). Finally, the resulting set of time series, $S_i' = \left(S_i'(1), \cdots, S_i'(L)\right)$, is used as teaching data for setting the node states of the VARDNN.

Performance Evaluation

We quantified the performance of the VARDNN-GC and VARDNN-DI measures over four different evaluations. The first evaluation confirmed the ability of VARDNN to learn and reproduce time series generated from uniformly distributed random numbers. The second evaluation estimated the ground truth directed graph from synthetic fMRI BOLD signals generated by the DCM, in order to clarify the relationship between network density and estimation accuracy. The third evaluation estimated the ground truth directed graph from synthetic neural activity signals generated by the reduced Wong-Wang model. This was done to clarify the relationship between node number and estimation accuracy. The DCM and reduced Wong-Wang models were chosen to generate biologically plausible synthetic neural signals for evaluation. The final evaluation was to measure VARDNN's performance on experimental data. Here, we evaluated Alzheimer's disease group separation ability using empirical rsfMRI BOLD signals. For each evaluation, we also compared our approach to other well-known functional connectome analysis algorithms, listed in **Table 1**. Of these algorithms, two are *zero-lag*-based algorithms: FC and PC. FC is the simplest measure of *pairwise* similarity between two time series, based on covariance. We used the corr() function from MATLAB (R2019b) for the comparison. PC is a conditional version of FC. It refers to the normalized correlation between two time series. We used the partialcorr() function of MATLAB to calculate the PC. Because PC relies on linear regression, it may face large dimensionality and multi-collinearity problems. Therefore, we implemented several countermeasure combined versions of PC, namely Principal Component PC (PC-PC), Elastic Net PC (EN-PC), and Partial Least Squares PC (PLS-PC). Elastic Net requires hyperparameters α and λ; we calculated these values from the minimum error of fivefold cross-validation using half of the sample data.

The *pairwise* GC and *multivariate* (or conditional) GC algorithms were implemented as MATLAB scripts. The GC time lag p parameter was empirically chosen as 3 for the evaluation. Multivariate GC, implemented as a multivariate vector auto-regression (mVAR) model, attempts to detect causal effects by considering differences between the network with all nodes and the network with the removal of a single node. On the other hand, the classical pairwise model simply considers two-node interactions. The mVAR model is widely used in many different fields, but it is also expected to suffer from the problems of data with large dimensionality and multicollinearity—the same as for the PC case. Therefore, we also implemented several counter-measure combined mvGC versions, namely multivariate Principal Component PC (mPC-GC), multivariate Elastic Net GC (mEN-GC), and multivariate Partial Least Squares GC (mPLS-GC) for our comparison.

Partially Conditioned GC (PCGC) is Granger causality based on partial conditioning, proposed by Marinazzo et al. (2012). The MATLAB script of its implementation was obtained from the GitHub.[1] Parameter "ndmax" of the init_partial_conditioning_par_m() function was calculated based on node number, and the partial_CGC_fix_nd_m() function was used to acquire the dFC matrix.

The RNN-GC implementation was obtained from the GitHub.[2] The evaluation environment was built using Anaconda3,[3] Python-3.6, Keras-1.2.2, and Tensorflow-1.14.0.[4] The dFC matrix was estimated as the average of 8 runs of the RNN_GC() function. In the first evaluation, we chose 30 as the number of hidden neurons.

Transfer entropy (TE) is a powerful tool for detecting the transfer of information between nodes and is capable of discovering non-linear interactions. It has been proven that TE is equivalent to Granger causality for data that can be assumed to be drawn from a Gaussian distribution (Barnett et al., 2009). We used the MuTE implementation (Montalto et al., 2014) in MATLAB for our comparison. This implementation automatically chooses the time lag based on the Bayesian information criterion. MuTE has three entropy estimators: the linear estimator (LIN), binning estimator (BIN), and nearest neighbor estimator (NN). Either a uniform or non-uniform embedding scheme is used to derive the transfer of information. As per the author's results, BIN-NUE shows good estimation in both the linear and non-linear cases, but our initial tests found that estimating a dFC matrix from the signals generated by DCM was unsuccessful. Therefore, we chose the LIN-UE and NN-NUE algorithms for our comparison.

In sections "VARDNN Performance With Respect to Network Density" and "VARDNN Performance With Respect to Node Count," various measures were used to estimate the ground truth directed graph, and in order to quantify them, evaluation was performed by changing the acceptance threshold and drawing an ROC curve. The ground truth matrix and result matrix were compared across the maximum value of the result matrix to its minimum. From this changing threshold, true and false positive rates of the result matrix were detected, and a ROC curve was

[1] https://github.com/danielemarinazzo/PartiallyConditionedGrangerCausality
[2] https://github.com/shaozhefeng/RNN-GC
[3] https://www.anaconda.com/
[4] https://www.tensorflow.org/

TABLE 1 | Comparison algorithms.

Name	Abbreviation	Type	Strategy	Implementation
Functional connectivity (Correlation)	FC	Zero-lag	Pairwise	MATLAB function
Partial correlation	PC	Zero-lag	Multivariate	MATLAB function
Principal component + Partial correlation	PC-PC	Zero-lag	Multivariate	Our Script (MATLAB)
Elastic net + Partial correlation	EN-PC	Zero-lag	Multivariate	Our Script (MATLAB)
Partial least square + Partial correlation	PLS-PC	Zero-lag	Multivariate	Our Script (MATLAB)
Pairwise Granger causality	pwGC	Predictive	Pairwise	Our Script (MATLAB)
Multivariate Granger causality	mvGC	Predictive	Multivariate	Our Script (MATLAB)
Principal component + Multivariate Granger causality	mPC-GC	Predictive	Multivariate	Our Script (MATLAB)
Elastic Net + Multivariate Granger causality	mEN-GC	Predictive	Multivariate	Our Script (MATLAB)
Partial Least Square + Multivariate Granger causality	mPLS-GC	Predictive	Multivariate	Our Script (MATLAB)
Partial Conditioning Granger causality	PCGC	Predictive	Multivariate	Script (MATLAB)
Recurrent neural network Granger causality	RNN-GC	Predictive	Multivariate	Script (Python)
Linear estimator-uniform embedding (Transfer entropy)	LINUE-TE	Predictive	Multivariate	MuTE (MATLAB)
Nearest neighbor estimator-Non-uniform embedding (Transfer Entropy)	NNNUE-TE	Predictive	Multivariate	MuTE (MATLAB)

generated from (0,0) to (1,1). After completing the ROC curve an area under the curve (AUC) value could be obtained. An AUC value close to 1 means a better classifier, whereas an AUC value close to 0.5 means that it is a random classifier. In this way, the accuracy of each measure was quantified by the value of AUC and used for our evaluation.

Synthetic Data for Algorithm Evaluation

Uniformly Distributed Random Data

In section "VARDNN Trained on Random Independent Signals," 8-nodes ($i = 1 \ldots 8$) of 100 length time series were generated by the Mersenne Twister (Matsumoto and Nishimura, 1998) pseudo random generator, with values in the range [0, 1] of the uniform distribution. The VARDNN was trained with these random signals to verify the DNN unit functionality of each node.

DCM Generated Data

In section "VARDNN Performance With Respect to Network Density," we evaluated the relationship between network density and ground truth network estimation from signals generated by an 8-node DCM model. Eight nodes were chosen based on the confirmation that the VARDNN operated as expected based on the results given in section "VARDNN Trained on Random Independent Signals." We designed 5 patterns of sparsely connected directed networks for generating synthetic fMRI BOLD signals with DCM. The network densities were 0.2, 0.25, 0.3, 0.41, and 0.5. Each DCM connectivity matrix $\boldsymbol{A}$ was derived from these network graphs. Each node had self-connections with a weight of 0.2, and inter-node connections had random weight values in the range [0.2, 0.5]. The DCM input connection matrix $\boldsymbol{C}$ was set to identity so that each node received endogenous fluctuations as random white noise to model the external environment. To generate synthetic fMRI BOLD signals, parameters of $T = 300$ (number of observations) and $TR = 2$ (repetition time) were used. Endogenous fluctuations for each node were generated by the spm_rand_mar() function, and fMRI BOLD signals were generated by the observer function spm_int_J() with spm_fx_fmri and spm_gx_fmri routines. In addition to the sparse network densities, we also evaluated networks with dense connectivity at the levels found in anatomical studies of the mammalian brain, for example 97% of all possible cortico-cortical connections have been observed in mouse anatomical results (Gămănuţ et al., 2018). Here, all nodes were connected (giving a network density of 1.14 when including self-connections), with uniform random weight values of $[-0.1, 0.1]$ for designating weak connections, and weight values of [0.2, 0.5] for strong connections. Due to the DCM formulation, we found that such fully connected networks easily exhibit divergence, and signals could reach a plateau value. By setting negative weights on some of the weak connections, a stable signal could be produced. We designed 5 patterns of directed networks with strong weight ($|a_{ij}| > 0.2$) network densities set to 0.05, 0.11, 0.16, 0.27, and 0.36. In total, 10 patterns of 8-node fMRI BOLD signals and exogenous signals were generated by DCM for the evaluation.

Reduced Wong-Wang Generated Data

In section "VARDNN Performance With Respect to Node Count," we evaluated the relationship between network size and ground truth network estimation from synthetic neural activity signals generated by the reduced Wong-Wang model implemented in The Virtual Brain (TVB) platform. We generated global coupling matrices, defined as $\boldsymbol{C}$, for the reduced Wong-Wang model by down-sampling Allen's mouse structural connectome matrix (Oh et al., 2014; Melozzi et al., 2017) that was constructed from a number of anterograde tracer injections. We chose the reduced Wong-Wang model along with structural connectome data to produce biologically plausible dynamics for analysis. We wanted to evaluate causal measures in a more biologically plausible setting; real structural connectivity data could act as a useful proxy for defining a realistic ground truth network for our analysis. Furthermore, the reduced Wong-Wang model was chosen to explore network size over the DCM model, as the latter was found to be too computationally expensive when a large number of nodes were used. Node counts were 16, 32, 48, 64, 80, and 98. Similar to the DCM case, all nodes were

TABLE 2 | Demographics of HC and AD subjects.

Variables	AD (n = 32)	HC (n = 41)	p-value
Male/Female	14/18	17/24	0.851
Mean Age (SD)	72.4 (2.2)	72.1 (2.6)	0.454

connected with weight values in [−0.05, 0.01], with a fraction of 0.15 connections assigned a strong weight value ($|c_{ij}| > 0.2$). The TVB framework was run using Anaconda3 and Python-3.6. We made small modifications to TVB's reduced Wong-Wang implementation to obtain appropriate signals for the evaluation. The "state_variable_boundaries" constant was bound to the range [0, 1], which could clamp activity signals, and we modified this value [−2.5, 2.5] in order to preserve causal relationships. Because the reduced Wong-Wang model has an internal noise factor, $\eta_i(t)$, we wanted to use this noise signal as an exogenous input for the VARDNN. Therefore, both the endogenous noise and neural activity signals were recorded and exported for evaluation. In summary, 6 patterns were used to generate 16–98 node sets of neural activity and internal noise signals. These were generated at a sampling frequency of 64 Hz, a global coupling $G = 1$ set to provide sufficient dynamics between regions, and a noise amplitude $\sigma = 0.014$.

Empirical Data Used to Evaluate Vector Auto-Regressive Deep Neural Network Performance

Alzheimer's Disease Neuroimaging Initiative Database

Data used in the preparation of this article were obtained from the *Alzheimer's Disease Neuroimaging Initiative* (ADNI) database.[5] The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). **Table 2** shows the demographics of healthy controls (HC) and AD subjects from the *ADNI2* subset of the database. The number of subjects with HC and AD were 41 and 32, respectively. Younger subjects with AD were preferred; HC of similar ages were selected.

Resting State fMRI Data Pre-processing

fMRI BOLD data from the *ADNI2* was scanned using a 3.0 Tesla Philips MRI scanner and saved in the DICOM format. The experimental settings of the scanner were: flip angle = 80, repetition time (TR) = 3,000 ms, echo time (TE) = 30 ms, pixel size = 3.3 × 3.3 mm, slice thickness = 3.3 mm, matrix size = 64 × 64 × 48, and frame length = 140. More information on the resting state parameters can be found on the ADNI website.

We converted the DICOM data to 4D NIfTI files using the dcm2nii tool (Li et al., 2016). Images were re-orientated so that the anatomical axial, sagittal, and coronal planes appeared correctly. Next, the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012) was used to pre-process the fMRI data. CONN registered NII images to the standard MNI brain space. Slice-timing correction was 1, 3, 5, . . ., 2, 4, and 6 for the Philips 3T MRI. We removed the first 10 frames of fMRI data, and the remaining data was smoothed by a FWHM of 5 mm. After pre-processing, the fMRI data was segmented using the CONN default atlas (132 ROIs), which is a combination of the FSL Harvard-Oxford atlas cortical and subcortical areas and the AAL atlas cerebellar areas. Voxels in each ROI were averaged to give representative fMRI BOLD signals for 132 regions over 130 frames.

Subject-Wise Evaluation Framework for Algorithm Comparison

We analyzed fMRI BOLD signals of HCs and ADs using 12 algorithms in **Table 1** except RNN-GC and NNNUE-TE to estimate connectivity between brain regions. This generated 132 × 132 sized functional connectome matrices for each of the 41 HC's and 32 AD's results. These two groups were compared using statistical testing. We performed the Lilliefors test to check the distribution of functional connectome matrices and the Wilcoxon rank sum test (Mann-Whitney U-test) to compare the two groups. From this comparison, we obtained a 132 × 132 p-value matrix describing the actual performance of lesion detection by each analysis algorithm. The top N_{rr} (= 30–300 step by 30) most different regional relationships between the HC and AD (that is, the set with the N_{rr} smallest p-values in the matrix) were then used to separate subjects to HC or AD group If an input functional connectome value (for a particular regional relationship) generated from new subject data was within the range $[-\beta\sigma, \beta\sigma]$ (where β = 1.5–2.0 with a stepsize of 0.1) of the AD's functional connectome population distribution, it is counted as "unhealthy"; otherwise, it is counted as "healthy." This was done for each of the input functional connectome values in the N_{rr} most different regional relationships and a "healthy count" was acquired. The "health rating" for a subject is simply derived by dividing the health count by N_{rr}.

$$\textit{Health rating} = \frac{\textit{Healthy count } \left(\textit{of } N_{rr} \textit{ regional relationships}\right)}{N_{rr}} \quad (17)$$

Two thresholds of N_{rr} and β were swept over for the subject-wise evaluation. From this the ROC curves and area under the curves (AUCs) of a total of 60 conditions were calculated to quantify and compare the performance of each of the studied analysis algorithms. Note that this subject-wise evaluation framework cannot be used as a reliable diagnostic technique if the subject number is not large enough—which we realize is the case for the dataset we used in this section. In such cases, cross-validation will show different p-value matrices and top N_{rr} regional relationships and give unstable diagnosis results. However, if there are enough subjects, we believe the presented statistical framework can be reliably applied for AD diagnosis.

[5] adni.loni.usc.edu

Statistical Analysis Methods

To test the obtained AUC values, we used Kruskal-Wallis ANOVA in MATLAB for significance testing, and the Steel test in R-4.0.3 for multiple comparisons. For normality testing of the functional connectome distribution, we used the Lilliefors test in MATLAB. To perform non-parametric testing between the HC and AD groups, we used the Wilcoxon rank sum test (Mann-Whitney U-test) in MATLAB. Statistical significance was set at $p < 0.05$.

RESULTS

Vector Auto-Regressive Deep Neural Network Trained on Random Independent Signals

Figure 2 shows the result of training the VARDNN on fully random and independent signals for 8-nodes. The following hyperparameters were used: training epochs = 1,000, and hidden layer 1 neuron count "hidden1" = 32 and layer 2 count "hidden2" = 22. The number of neurons per layer were empirically calculated by the method described in **Supplementary Figure 1**. Before training, the VARDNN nodes output signals clearly different from the teacher data, but after training with the conditions depicted in **Figure 2A**, the VARDNN nodes were able to output signals similar to the teaching signal (**Figure 2**). Before training, the mean absolute error (MAE) was 0.456 and after training it became 0.018. **Figures 2C,D** show some performance measures for VARDNN training. The MAE and training time were measured while changing the number of nodes and signal length, with the number of training epochs = 1,000. For any node number the VARDNN converged well and MAE showed values less than 0.02. Training time took less than 1 min with 30 nodes and signals of length 100 frames, and around 9 min with 180 nodes and signals of length 1,000 frames on a computer with the following specifications: OS, Windows 10 Pro; CPU, AMD Ryzen Threadripper 3970X 32-Core Processor (3.70 GHz); Memory, 128 GB. This makes VARDNN training very practical for data needing 100 nodes or a similar order of magnitude (such is the case when considering the number of cortical regions defined in brain atlases). Next, we performed a simple check to confirm VARDNN's ability to estimate dFC; node 6′s signal was copied to node 2 and 4 at the next time step. The VARDNN was trained with new network (**Figure 2E**). Zero-lag analysis (FC) detected a strong relationship between nodes 2 and 4 (**Figure 2F**).

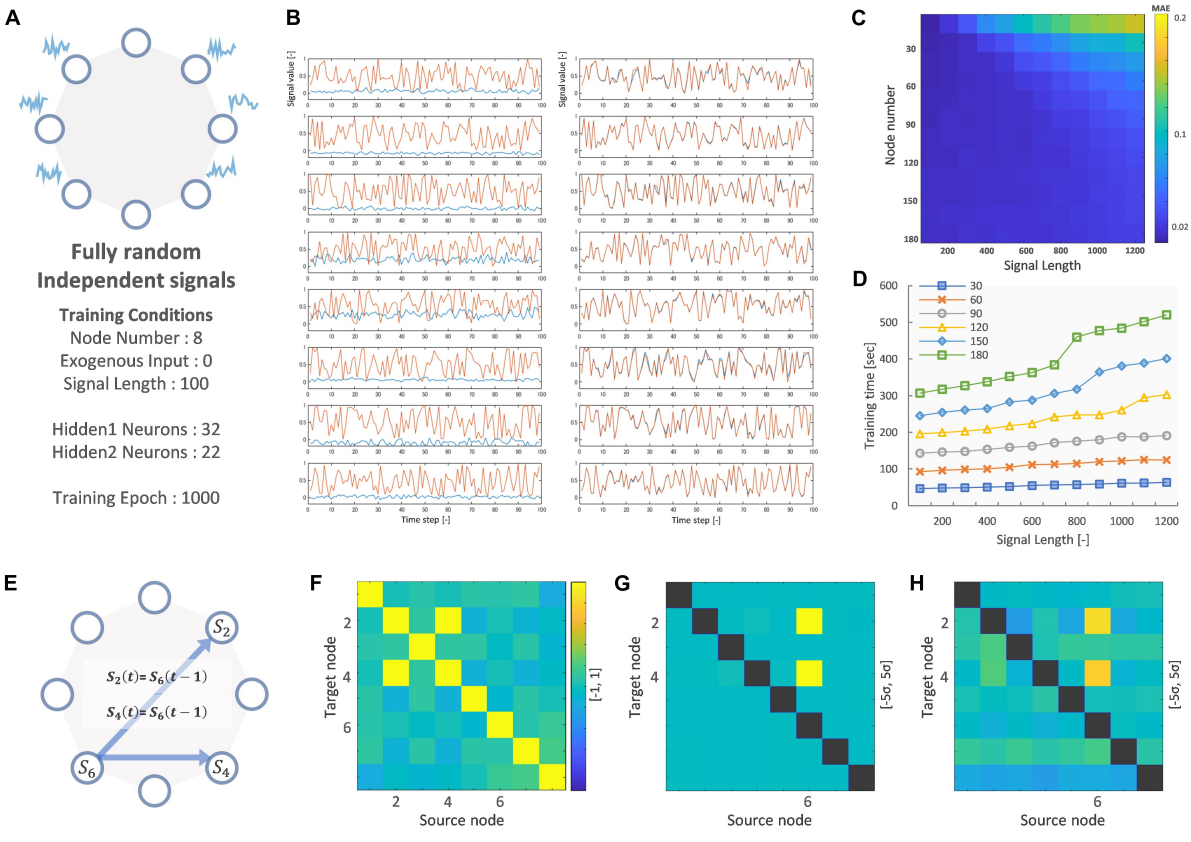


FIGURE 2 | (A) Node network image and training conditions. **(B)** Left shows before training signals of 8 nodes (Red: teaching signal, Blue: VARDNN output). Right shows after training signals. **(C)** Training error result (MAE) of signal length and node count. **(D)** Training time result of signal length and node count. **(E)** Node network image for evaluating FC, mvGC and VARDNN-GC. **(F)** FC result of E. Color range is [–1, 1]. **(G)** mvGC result of E. Color range is *[–5σ, 5σ]*. **(H)** VARDNN-GC result of E. Color range is *[–5σ, 5σ]*.

On the other hand, mvGC and VARDNN-GC detected strong causality from node 6 to nodes 2 and 4 (**Figures 2G,H**). Since Granger causality returns values that differ by two or more digits, the matrix is displayed as a deviation.

Vector Auto-Regressive Deep Neural Network Performance With Respect to Network Density

Figures 3, **4** A show evaluation results for the relationship between sparse pattern network density and ground truth estimation from signals generated by the 8-node DCM. The signal length was 300, and the trial number was $N = 8$. The VARDNN was trained with the synthetic BOLD signals and exogenous signals with training epochs = 1,000, hidden1 neurons = 55, and hidden2 = 37 (determined by the method described in **Supplementary Figure 1**). Both VARDNN-DI and VARDNN-GC show competitive results against conventional algorithms. In particular, VARDNN-DI (AUC: 0.92–0.94 across densities) shows a significantly higher AUC score compared to other algorithms, or a similar AUC score to mEN-GC (0.99–0.89), LINUE-TE (0.93–0.87), and higher as the density of strong connections increased. The zero-lag analysis algorithms, FC and PC families, showed similar results in all cases. mPLS-GC, mPC-GC, PCGC, RNN-GC, and NNNUE-TE sometimes showed a high AUC score but appeared unstable. mEN-GC, LINUE-TE, and mvGC showed higher AUC scores relative to the others, except for the VARDNN-based algorithms (especially in the high network density case). **Figures 4B**, **5**, show evaluation results for the relationship between strong weight ($|a_{ij}| > 0.2$) fully connected network density and ground truth estimation. Once again, VARDNN-based algorithms also showed competitive results against existing algorithms. As in the sparse network evaluation, VARDNN-DI (AUC: 1.0–0.94) showed a significantly higher AUC score over most algorithms and a similar AUC score to mEN-GC (1.0–0.92), LINUE-TE (1.0–0.92), and became higher as the density of strong connections increased. Increased network density causes synergistic signal activation and multicollinearity problems and VARDNN demonstrated robust capabilities for dealing with these issues.

Computation time depends not only on the efficiency of the algorithm, but also on its implementation and runtime environment. However, as a basic guide we provide the computation times of the implementations we tested. The machine specification was the same as that described in section "VARDNN Trained on Random Independent Signals." Many algorithms, such as FC, PC, mvGC, pwGC showed an order of about 10^{-2} s for 8 nodes. LINUE-TE (0.05 s) showed an order of about 10^{-1} s. VARDNN-based algorithms (24 s) and NNNUE-TE (8 s) showed an order of about 10 s. On the other hand, RNN-GC (599 s) showed an order of more than 10^2 s. Because of this, RNN-GC was excluded from the evaluation in section "VARDNN Performance With Respect to Node Count" due to an excessively long calculation time in the large node case. In addition, NNNUE-TE was also excluded from the same evaluation due to a memory usage error.

Vector Auto-Regressive Deep Neural Network Performance With Respect to Node Count

Figures 6, **7** show evaluation results for the relationship between node count and ground truth estimation of node signals generated by the reduced Wong-Wang model. Conventional algorithms such as multivariate GC (AUC: 0.79–0.5), pairwise GC (AUC: 0.82–0.53), and LINUE-TE (AUC: 0.83–0.54) struggled when the number of nodes was large. As they are based on OLS, their sensitivity is reduced by the number of free parameters. Countermeasure combined mvGC showed a small advantage over standard mvGC. On the other hand, VARDNN-DI (AUC: 0.88–0.64) and VARDNN-GC (AUC: 0.74–0.66) showed better scaling performance. At node numbers above 48, VARDNN-DI and VARDNN-GC showed significantly higher AUC scores over LINUE-TE and the mvGC family, and above 64 was significantly higher than pwGC. The pwGC algorithm showed better scores than mvGC; this may be due to the number of free parameters. VARDNN-GC did not perform well for the smallest network size, but unlike the rankings in section "VARDNN Performance With Respect to Network Density," it outperformed VARDNN-DI for more than 64 nodes. Many zero-lag analysis methods did not perform well in any node count. This was because the sampling frequency (64 Hz) was too fast and correlations between nodes were lower than the fMRI BOLD signal repeat time (0.5 Hz). Overall, VARDNN showed a better AUC over linear regression results. This means VARDNN has the ability to be robust to redundancy or dealing with large dimensionality data. Interestingly, VARDNN shows a higher AUC over countermeasure techniques. This difference may be due to VARDNN's non-linearity.

Vector Auto-Regressive Deep Neural Network-Directional Influence and -Granger Causality Can Successfully Analyze Alzheimer's Disease

The rsfMRI data acquired from *ADNI2* was preprocessed then separated by the CONN default atlas, generating 132 regions and 130 frames of rsfMRI BOLD signals. These were analyzed by several analysis algorithms to obtain functional connectome matrices. VARDNN was trained with the empirical BOLD signals and randomly generated exogenous signals (uniform distribution in range [0, 1], acting as a noise factor to the algorithm), with training epochs = 1,000, hidden1 neurons = 34, and hidden2 = 23 (calculated by the method described in **Supplementary Figure 1**). **Figures 8A–D** show the mean functional connectome matrices for the AD and HC subjects. Most algorithms showed similar patterns for AD and HC, but PC, mvGC, and LINUE-TE did not (**Figure 8E**). The cosine similarities between AD and HC matrices were FC (0.965), PLS-PC (0.645), pwGC (0.976), VARDNN-DI (0.976), PC (0.11), mvGC (−0.025), and LINUE-TE (0.034). This meant that FC, pwGC, VARDNN, countermeasure combined PC and mvGC consistently captured meaningful regional relationships across the AD and HC subjects,

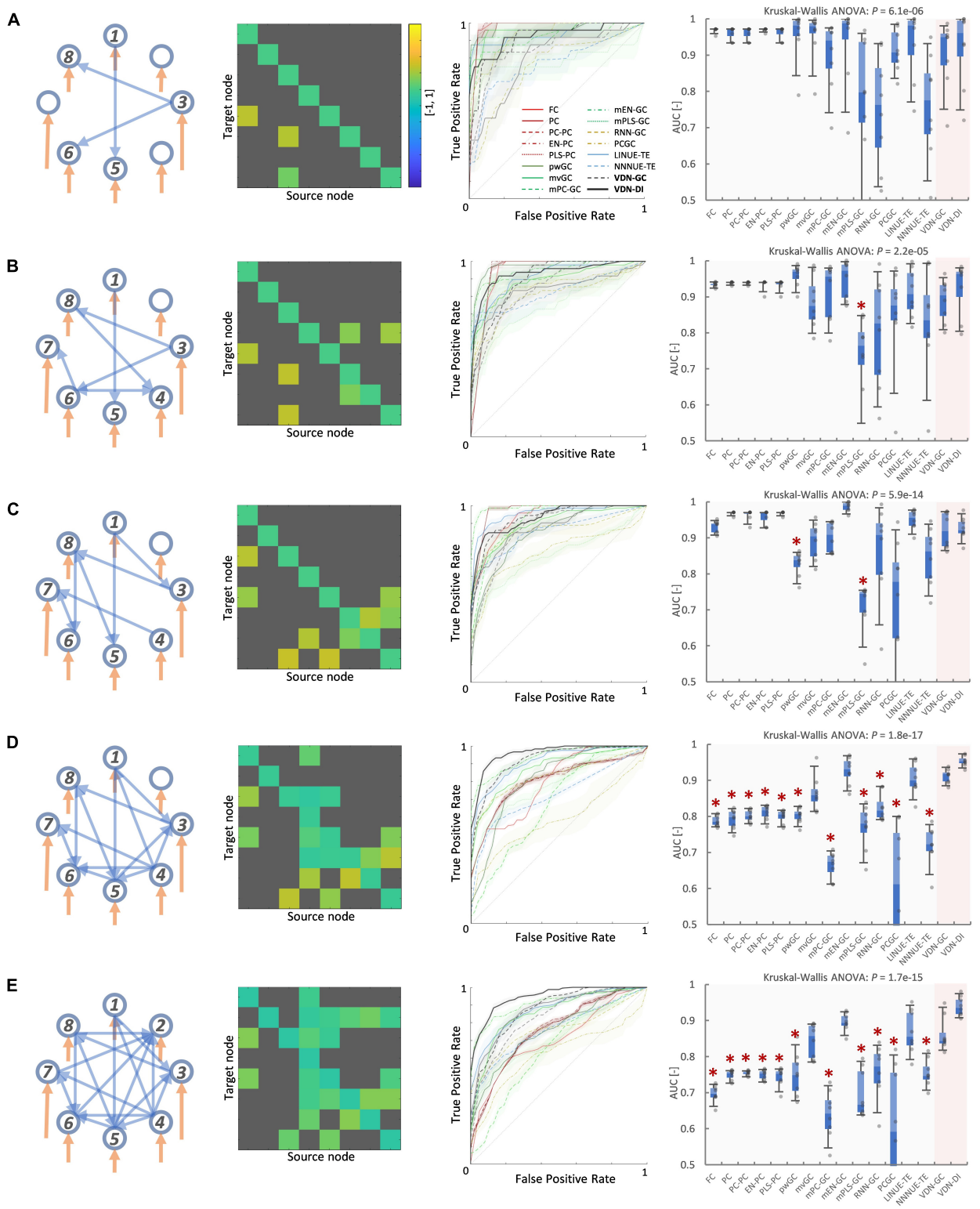


FIGURE 3 | (A) Sparse pattern of 8-node network of network density 0.2. Eight self-connections and three other connections. From left to right images show the ground truth network graph, the matrix *A* for the DCM, the ROC curve result of ground truth estimation by each analysis algorithms, and the box-plot of AUC results, respectively. The color bar describes the matrix *A*-values for **(A–E)**. In the box-plot, a black dot shows each experimental result and a red asterisk shows significant difference from VARDNN-DI to other algorithms by Steel test (*$p < 0.05$) ($N = 8$). **(B)** Result of network density 0.25 (sparse pattern of 8-node network). **(C)** Result of network density 0.3. **(D)** Result of network density 0.41 **(E)** Result of network density 0.5.

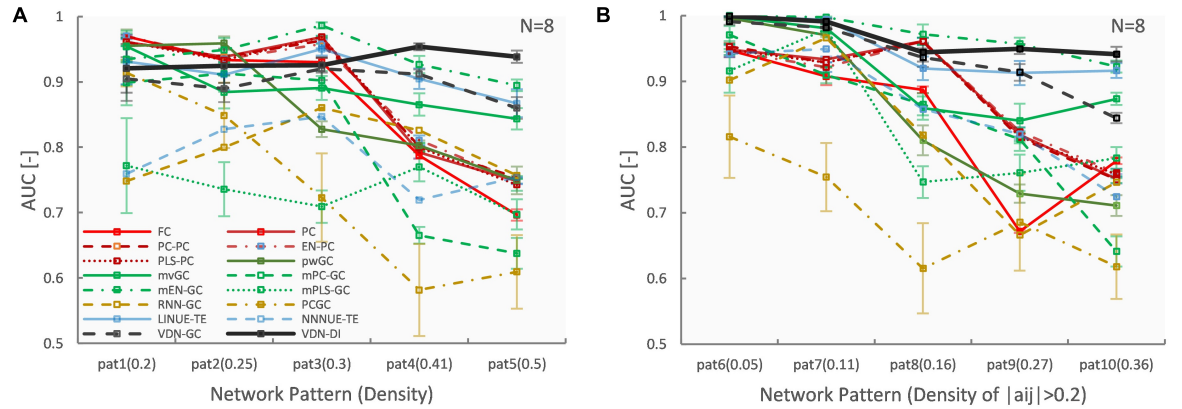


FIGURE 4 | (A) Network density (sparse pattern) vs. AUC result. Error bar shows standard error. **(B)** Network density (fully connected) vs. AUC result.

but PC, mvGC, and LINUE-TE were clearly insensitive because of the large dimensionality problem.

To compare the values of functional connectome matrices (132 × 132) between the AD and HC groups we needed to test the normality of their distributions. FC values in the matrices for the HC group (21.3%) were significantly *not* Gaussian distributed ($p < 0.05$) (**Figure 8F**, left). VARDNN-DI values in the matrices for the HC group (74.1%) were significantly *not* Gaussian distributed ($p < 0.05$) (**Figure 8H** right). From these results, a non-parametric test was chosen as appropriate for comparing the AD and HC groups. We used the Wilcoxon rank sum test (Mann-Whitney *U*-test) for this comparison.

Figure 9 shows comparison results between the AD and HC groups for each analysis algorithm. VARDNN-DI detected that 12% of regional relationships were significantly affected ($p < 0.05$) by Alzheimer's disease; FC (10.5%), PLS-PC (5.5%), and pwGC (7.3%) were also able to detect a difference well. Even though these algorithms had functional connectome matrices showing high cosine similarity between AD and HC (**Figure 8E**), they could capture significant differences between the AD and HC groups. Next, we confirmed the consistency of our results with previous studies. **Table 3** shows the top 20 most different regions (in terms of dFC) between the AD and HC groups, calculated from VARDNN-DI.[6] Significantly different dFC were counted by brain region, and the total number of both sources and targets is displayed in the VDNdiTotal column in descending order. **Table 3** columns are sorted by the bold values in the VDNdiTotal column. In this result, VARDNN-DI was able to detect lesion effects of the Default Mode Network (DMN), including precuneus cortex, angular gyrus, and cingulate gyrus (posterior division). Additionally, supramarginal gyrus (posterior division) and intracalcarine cortex were also detected and matched well with previous studies. This result confirms that whole-brain screening with VARDNN-DI may be useful for further brain analysis. Both FC and pwGC detected some common effects within the DMN. Although these causal and zero-lag analyses use different approaches, they were capable of showing similar results.

Figure 10 shows the result of Alzheimer's disease separation ability by our subject-wise evaluation framework. We tried to use two subject groups as a ground truth to quantify the separation ability of each algorithm. **Figure 10A** shows the top N_{rr} (= 100) most different regional relationships between the AD and HC groups by VARDNN-DI. From left to right, regional relationships were sorted in ascending order of *p*-value. As shown in this graph, the dFC distribution of AD was significantly different from that of HC. This means that connectivity is reduced in these regions. Then, we were able to use these N_{rr} regional relationships to separate AD and HC subjects. The functional connectome matrix was calculated from the rsfMRI BOLD signal obtained from a new subject, and the top N_{rr} regional relationships from the matrix were used to obtain their *health rating* (**Figure 10B**), which could be used to distinguish between AD (N = 32) and HC subjects (N = 41) (**Figure 10C**). ROC curves were generated from the subject distribution of AD and HC based on the *health rating* histogram. A "healthy" threshold to classify healthy or unhealthy was moved from 100 to 0% over the *health rating* range. In total, 60 conditions of the framework settings (N_{rr}:30–300 with step-size 30, β threshold: 1.5–2.0 with step-size 0.1) were tested for each algorithm. **Figure 10D** shows ROC curve results for each analysis algorithm, and **Figure 10E** shows the box-whisker plot of AUCs. VARDNN-DI (0.97) showed the highest AUC and PLS-PC (0.969) showed a competitive AUC result. These algorithms separated the AD and HC groups well. FC (0.897) and pwGC (0.888) (both pairwise strategies) showed similar cosine similarities and subject-wise evaluation results in TR = 3 s rsfMRI BOLD data. They in turn showed a similar ability for group separation. The mvGC family of algorithms and LINUE-TE showed unstable subject-wise evaluation results—they did not separate the two groups well. The PC family showed high AUC scores, but they were unstable over several thresholds. Also, their cosine similarity showed unstable results, i.e., their similarities were smaller than other algorithms (**Figure 8E**) and significantly different regional relations were fewer (**Figure 9**). For many brain

[6]VARDNN-DI was chosen over VARDNN-GC as we found it gave better diagnosis results, implying that VARDNN-GC's estimated significantly different regions were less accurate.

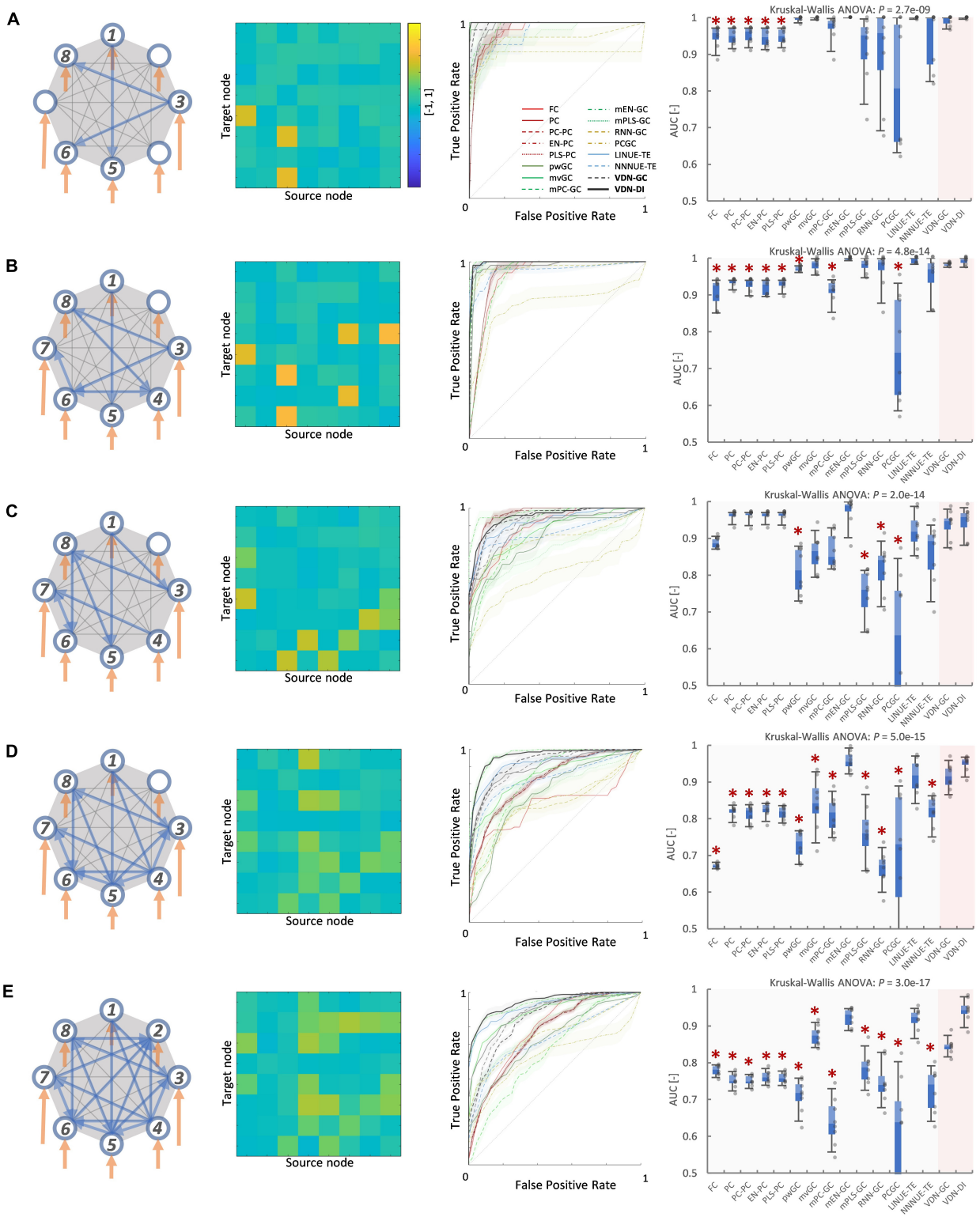


FIGURE 5 | (A) Fully connected 8-node network with strong weight (| aij| > 0.2) network density 0.05. From left to right, images show the ground truth network graph, the matrix *A* for DCM, the ROC curve result of ground truth estimation by each analysis algorithm, and the box-plot of AUC results, respectively. The color bar describes the matrix *A*-values for **(A–E)**. In the box-plot, a black dot shows each experimental result and a red asterisk shows significant difference from VARDNN-DI to other algorithms by Steel test (*$p < 0.05$) ($N = 8$). **(B)** Result for the strong weight (| aij| > 0.2) network density 0.11 (fully connected 8-node network). **(C)** Result of strong weight network density 0.16. **(D)** Result of strong weight network density 0.27. **(E)** Result of strong weight network density 0.36.

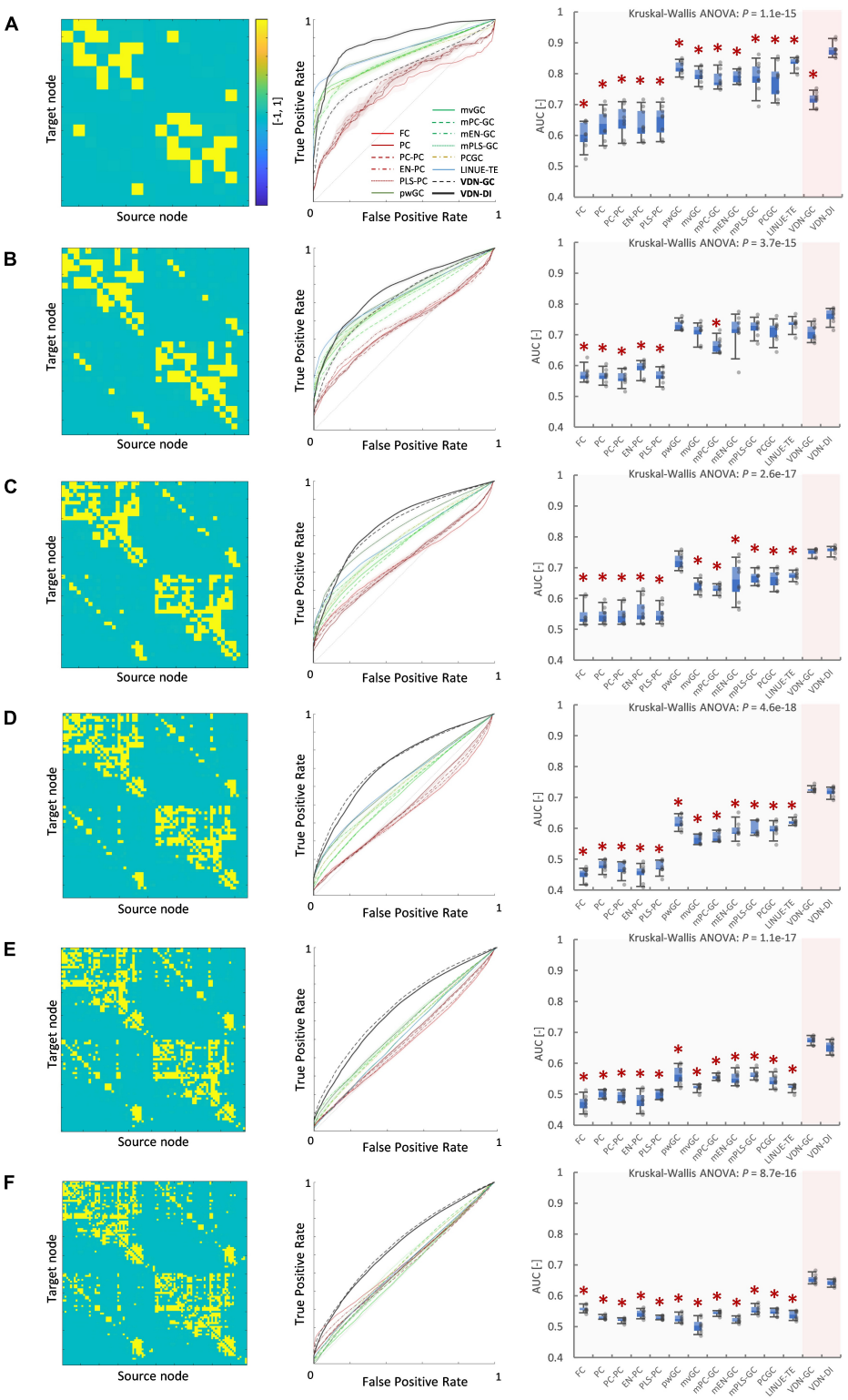


FIGURE 6 | (A) Fully connected 16-node network with strong weight (| cij| $>$ 1) network density 0.15. Node signals were generated by the reduced Wong-Wang model from the TVB software. From left to right images show the global coupling matrix *C* for the reduced Wong-Wang model, the ROC curve result of ground truth estimation by each analysis algorithm, and the box-plot of AUC results, respectively. Color bar shows matrix *C*-value for **(A–F)**. In the box-plot, black dot shows each experimental result and red asterisks denote a significant difference from VARDNN-DI to other algorithms by the Steel test (*$p < 0.05$) ($N = 8$). **(B–F)** Results for the 32-node, 48-node, 64-node, 80 node, and 98-node networks, respectively.

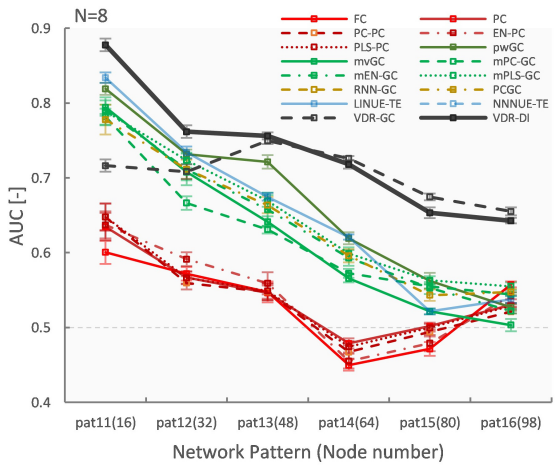


FIGURE 7 | Node number (fully connected) vs. AUC result.

regions, the conditioning strategy shows very small differences around the zero value and the two group distributions were mixed. Even though some PC approaches were able to show good group separability, overall results were unstable. In summary, this demonstrates that VARDNN performs well with real-world, empirical rsfMRI BOLD data.

DISCUSSION

In this article, we presented VARDNN, a novel method for functional connectome estimation based on a vector auto-regressive deep neural network architecture. We proposed two VARDNN dFC measures that show good sensitivity to causal structures, namely VARDNN-GC and VARDNN-DI. In our experiments, the VARDNN was successfully trained with random uniformly distributed signals, synthetic fMRI BOLD signals that were generated by the DCM model, synthetic neural activity signals that were generated by the reduced Wong-Wang model, and empirical fMRI BOLD signals acquired from the *ADNI2* dataset. Functional connectome estimation of ground-truth synthetic fMRI BOLD signals from DCM was successful, and VARDNN-DI and VARDNN-GC showed competitive results with existing functional connectome estimation measures. In particular, VARDNN-DI showed a higher AUC score over other algorithms for the high network density case (**Figure 5**).

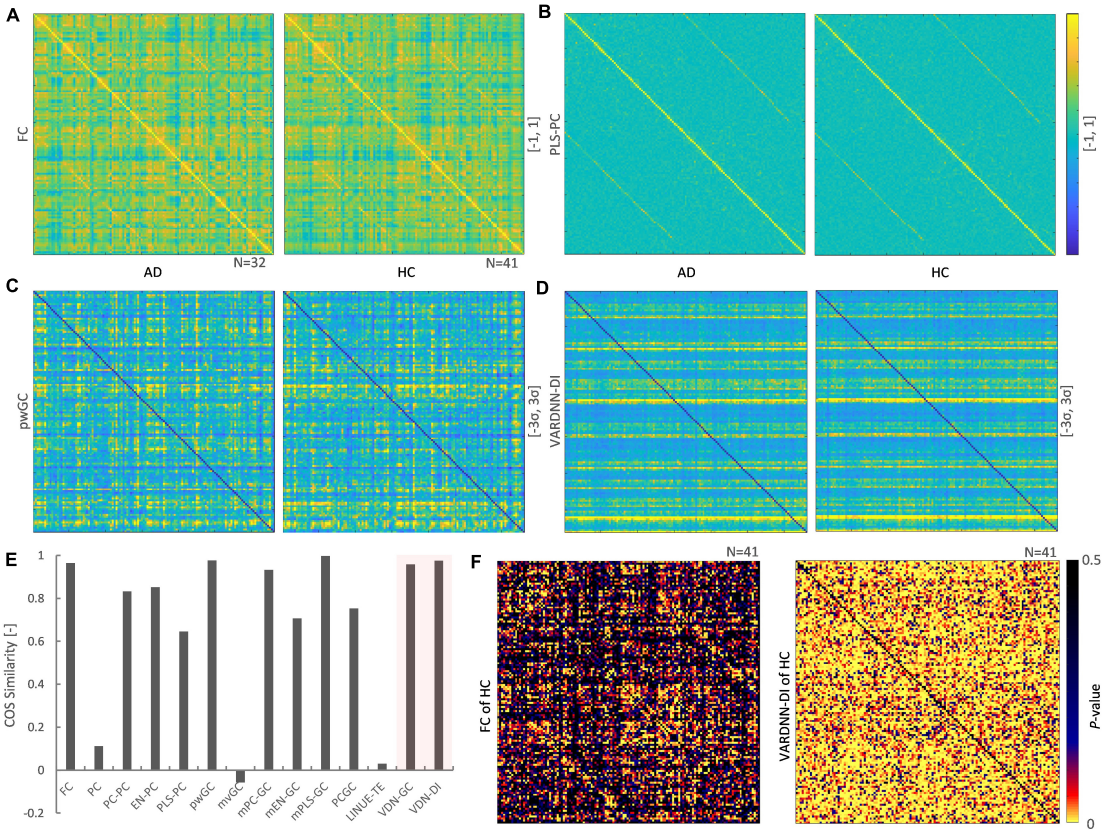


FIGURE 8 | (A) Left: mean FC matrix (132 × 132) of the AD group (*N* = 32). Right: mean FC matrix of the HC group (*N* = 41) Color range is [–1, 1]. Source is column, target is row. **(B)** Mean PLS-PC matrix (left: AD group, right: HC group). Color range is [–1, 1] **(C)** Mean pwGC matrix (left: AD group, right: HC group). Color range is [–3σ, 3σ] **(D)** Mean VARDNN-DI matrix (left: AD group, right: HC group). Color range is [–3σ, 3σ] **(E)** Cosine similarity of the mean matrix between the AD and HC group matrices. **(F)** Lilliefors test result matrix of FC (left) and VARDNN-DI (right) of HC.

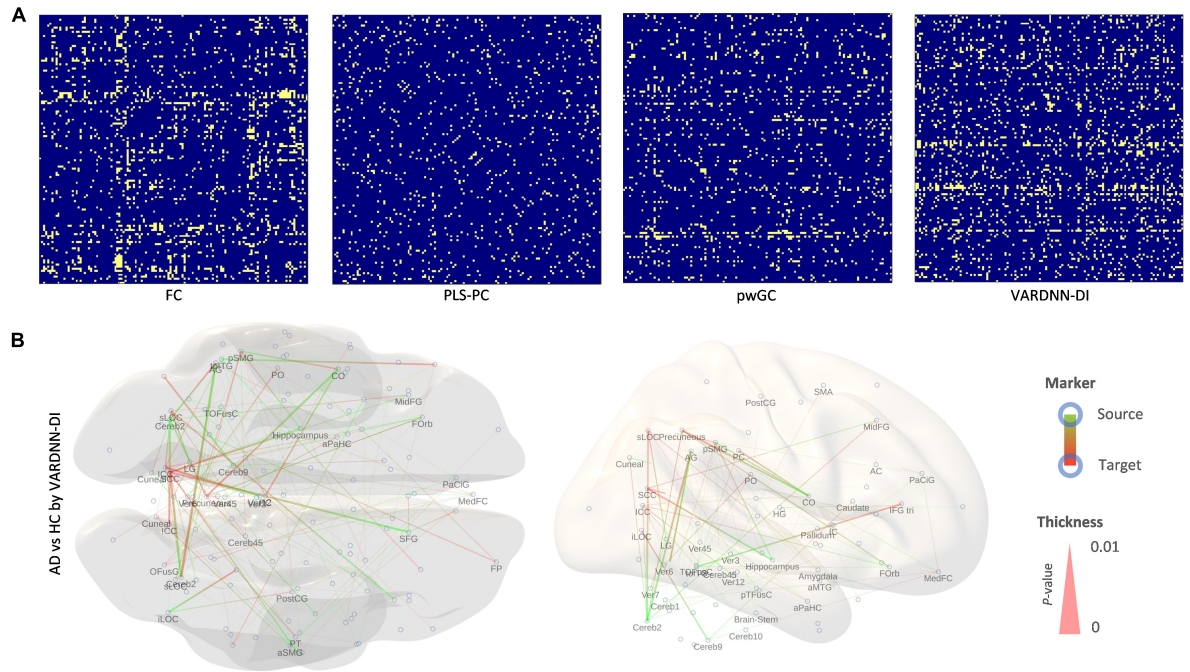


FIGURE 9 | (A) Significantly different ($p < 0.05$, yellow color; non-significant, blue color) regional relationship matrices between AD and HC groups. From left to right, FC, PLS-PC, pwGC, and VARDNN-DI, respectively. **(B)** The whole brain image of the most different ($0 \leq p < 0.0024$) regional relationships by VARDNN-DI. The image on the left is seen from the dorsal view and the right image shows the medial view of the left hemisphere. Posterior is left and anterior is right. Blue circles are region centroids. The source to target areas are color coded from green to red. A smaller p-value shows a thicker line.

Furthermore, functional connectome estimation from ground-truth synthetic neural activity signals generated by the reduced Wong–Wang model was evaluated. VARDNN-based dFC measures significantly outperformed other causal estimation measures for the large node count case (**Figure 6**). Several articles evaluating undirected or directed functional connectivity algorithms based on synthetic fMRI BOLD signals generated by DCM exist, for example, Smith et al. (2011) and Prando et al. (2020). However, the results of these experiments are difficult to compare due to their various network configurations and densities. For example, networks with node numbers of 5, 10, and 50 with very sparse connections were used in Smith et al. (2011), while node numbers of 7 and 66, with a network density 0.5 were used in Prando et al. (2020). In this work we evaluated both increasing network density and node number cases. The evaluation of multiple types of algorithms for the early diagnosis of Alzheimer's disease has been performed by Dauwels et al. (2010). We designed a subject-wise evaluation framework to separate AD and HC groups from empirical rsfMRI BOLD data in order to quantify the performance of functional connectome estimation measures. In our results, VARDNN-DI especially succeeded in extracting lesioned brain regions consistent with previous studies (**Table 3**), and it showed a higher AUC score over existing functional connectome estimation measures when separating AD and HC groups (**Figure 10**). This result also supports the sensitivity of the VARDNN-DI measure.

Both multivariate Granger causality and LINUE transfer entropy rely on linear multivariate auto-regression. Thus, their sensitivity becomes low for large node counts and little sample data (Barnett and Seth, 2014). Several countermeasure techniques, such as PCA (Zhou et al., 2009), Elastic Net (Zou and Hastie, 2005; Ryali et al., 2012), and PLS (Wold et al., 2001; Krishnan et al., 2011) have been used to avoid these large dimensionality and multi-collinearity problems. We compared VARDNN against these techniques and VARDNN showed competitive or even better estimation results. If a DNN does not have (non-linear) activation functions, it is essentially the same as linear regression. Activation functions, such as ReLU, non-linearly extend the ability of linear regression. This strategy works well for extending the capability of GC by VARDNN. The well-studied training techniques of deep neural networks complements the weaknesses of linear regression and extends its capabilities.

We showed that VARDNN-based GC was able to obtain better results than RNN-GC, an established neural network-based approach. RNN-GC was directly compared with another neural network-based approach called NN-GC, which it outperformed (Montalto et al., 2015; Wang et al., 2018). The original NN-GC definition of causality was a simple subtraction: $\boldsymbol{Err}_{reduced} - \boldsymbol{Err}_{full}$, and this definition showed less sensitivity than the RNN-GC definition of $\log\left(var\left(\boldsymbol{Err}_{reduced}\right)/var\left(\boldsymbol{Err}_{full}\right)\right)$. Our VARDNN-GC measure used the same type of definition as RNN-GC and outperformed it. From the comparative results of Wang et al. (2018) we can also expect VARDNN to outperform NN-GC. RNNs require sequential input, which can make training convergence difficult. Our approach uses an auto-regressive DNN

TABLE 3 | Top 20 regions of most different dFC between AD and HC, calculated from VARDNN-DI.

ROI name	Network	FC	PC-PC	PLS-PC	pGCSrc	pGCTrg	VDNdiSrc	VDNdiTrg	VDNdiTotal	References
Precuneous (Precuneous cortex)	pDMN	20	14	8	11	18	30	59	**89**	Bakkour et al., 2013; Peraza et al., 2014; Utevsky et al., 2014; Zott et al., 2018; Pascoal et al., 2019
pSMG l (Supramarginal gyrus, posterior division left)	FPN/Lang (L)	24	13	5	12	4	37	43	**80**	Harasty et al., 1999; Bakkour et al., 2013; Peraza et al., 2014
ICC l (Intracalcarine cortex left)	Visual.Primary	6	12	9	10	4	19	58	**77**	Peraza et al., 2014; Utevsky et al., 2014; Hafkemeijer et al., 2015; Schwab et al., 2020
SCC l (Supracalcarine cortex left)	Visual.Primary	14	8	13	10	12	10	66	**76**	Utevsky et al., 2014; Hafkemeijer et al., 2015
AG l (Angular gyrus left)	lDMN	13	16	6	8	26	18	58	**76**	Bakkour et al., 2013
PC (Cingulate gyrus, posterior division)	pDMN	5	9	8	19	10	35	24	**59**	Bakkour et al., 2013
toITG l (Inferior temporal gyrus, temporooccipital part left)	DAN	17	9	6	8	1	18	41	**59**	Mosconi et al., 2009; Bakkour et al., 2013; Peraza et al., 2014; Utevsky et al., 2014; Zott et al., 2018; Pascoal et al., 2019
CO l (Central opercular cortex left)	Auditory	9	9	5	6	7	35	23	**58**	
Cuneal r (Cuneal cortex right)	Visual.Primary	6	9	7	9	3	14	41	**55**	Utevsky et al., 2014; Hafkemeijer et al., 2015
Cereb6 l (Cerebelum 6 left)	Cerebellum	10	2	7	15	15	33	19	**52**	Jacobs et al., 2018
SFG r (Superior frontal gyrus right)	FPN (R)	22	15	12	12	11	37	14	**51**	Bakkour et al., 2013
Pallidum l	Salience	13	9	7	16	8	18	32	**50**	Bakkour et al., 2013
SPL r (Superior parietal lobule right)	DAN	36	12	5	12	6	34	16	**50**	Peraza et al., 2014
Cereb9 l (Cerebelum 9 left)	Cerebellum/pPaHC	10	11	7	7	8	42	7	**49**	Jacobs et al., 2018
FOrb l (Frontal orbital cortex left)	Inferior Temporal	3	7	4	4	13	37	12	**49**	
Caudate r	Thalamus	16	6	8	5	35	28	20	**48**	
AG r (Angular gyrus right)	lDMN	15	3	9	10	0	23	24	**47**	
Cereb45 l (Cerebelum 4 5 Left)	Cerebellum	9	10	5	12	22	16	31	**47**	Peraza et al., 2014
sLOC l (Lateral occipital cortex, superior division Left)	sLOC	17	4	5	4	19	28	19	**47**	Bakkour et al., 2013
Ver12 (Vermis 1 2)	Cerebellum/pPaHC	34	9	6	7	3	19	27	**46**	

From left to right column: ROI name, brain network name, significantly different region counted by FC, PC-PC, PLS-PC, pwGC source count, pwGC target count, VARDNN-DI source count, VARDNN-DI target count, total of VDNdiSrc and VDNdiTrg column, references, respectively.

structure so mini-batch and shuffling during network training becomes possible. This flexibility improved the training time of the VARDNN and the accuracy of causality estimation. Furthermore, we defined a new dFC measure, VARDNN-DI, as $\left|z_{full} - z_{reduced}\right|$, to extract causal relationships from trained DNN weights. This simple equation showed surprisingly high sensitivity in comparison to other algorithms. Owing to these properties, the VARDNN can be trained on non-linear signals and capture non-linear relationships in the weights of the DNNs.

Despite its effectiveness, there are some limitations to the VARDNN and its functional connectome measurements. Since the designs of the VARDNN-based dFC measures have similarities with *predictive*-based analysis, they have similar limitations to Granger causality (Bressler and Seth, 2011; Barnett and Seth, 2014; Stokes and Purdon, 2017). The first is variability in the shape and latency of Hemodynamic Response Functions (HRFs) in different brain regions (Miezin et al., 2000) and across subjects (Aguirre et al., 1998). Neural activity and HRFs in brain regions are not completely synchronized, so data driven functional connectome measures, including FC, PC, and GC, have always faced this issue. However, many biological experiments involve two group comparisons, for example AD vs. HC, wild type vs. mutated disease model, etc. In this scheme, the variability of HRFs in different brain regions and subjects can be considered as differentiating features between the two groups. Therefore, for such schemes, applying data driven functional connectome measures for rsfMRI BOLD signal analysis and striving to increasing the experimental number will support

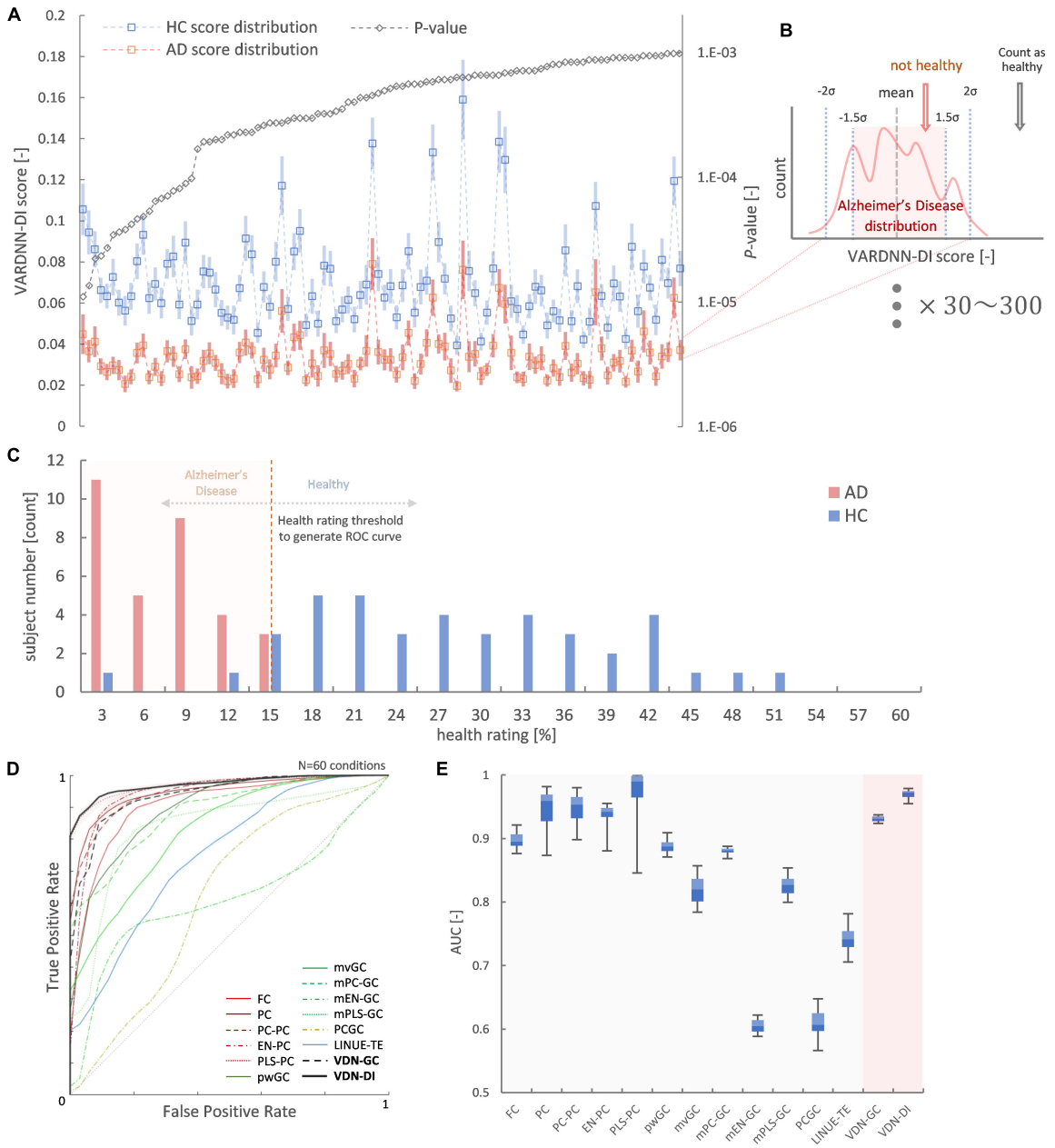


FIGURE 10 | (A) Top 100 most different regional relationships by VARDNN-DI. The left vertical axis shows VARDNN-DI score and blue (HC) and orange (AD) lines show mean score (box) and standard error (bar). The right vertical axis shows the logarithmic scale of *p*-value. The black dotted line shows the *p*-value of the regional relationship between AD and HC groups. **(B)** Healthy count estimation for subject-wise evaluation framework between AD vs. HC. If the input VARDNN-DI score is within the specific σ range (1.5–2.0) of the AD distribution, it is counted as unhealthy. This operation was repeated from top N_{rr} (30–300) most different regional relationships. **(C)** Example histogram (N_{rr} = 100, β = 2, VARDNN-DI) of *health rating*. Red bars show AD subjects. Blue bars show HC subjects. **(D)** Mean ROC curves of 60 conditions by each analysis algorithm. **(E)** AUC results from **(D)**.

robustness of experimental results. The second issue is related to changes in the sampling time interval. In our experiments, VARDNN-DI detected lesioned brain regions from experimental rsfMRI BOLD signals with a repetition time (TR) of 3 s. These dFCs clearly depend on the repetition time. If the repetition time was changed to 1.5 s, the dFC results might have differed. This problem is not only for *predictive*-based analysis, but also for *zero-lag* analysis. Longer time intervals average over fine patterns in neural activity, and correlation-based analysis shows better sensitivity at lower as compared to higher sampling rates. Thus, we need to be careful and consider this time dependence when choosing the appropriate analysis algorithm. The third is with regard to negative causal relationships between nodes. The DCM inversion algorithm is able to estimate positive and negative

connection weights, but GC-based algorithms basically do not have this function because they are calculated from changes in variance. Some GC approaches may generate negative values, but such values have no physical interpretation (Chen et al., 2006). This type of positive and negative causality estimation is also an issue for VARDNN.

We provide a VARDNN toolbox for MATLAB that implements the algorithms described in this paper. The code is open source and can be downloaded from the GitHub repository at https://github.com/takuto-okuno-riken/vardnn. The VARDNN toolbox contains a command line tool that includes several analysis algorithms, namely VARDNN-DI, VARDNN-GC, mvGC, pwGC, linear TE, FC, and PC to estimate a functional connectome from brain signals. We hope that this toolbox will help further brain analysis, and the toolbox will be expanded with new features in the future. The Deep Learning toolbox ver. 12.1 (or later), Fuzzy Logic Toolbox ver. 2.6 (or later), and MATLAB R2019a (or later) are required to run the VARDNN scripts.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Alzheimer's Disease Neuroimaging Initiative. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TO contributed to conception and design of the study and wrote the first draft of the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

This research was supported by the program for Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) from the Japan Agency for Medical Research and Development, AMED. Grant number: JP20dm0207001. Data collection and sharing was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health, grant number: U01 AG024904) and DOD (Department of Defense, award no. W81XWH-12-2-0012). ADNI was funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC.; Johnson and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research was providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data were disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.764796/full#supplementary-material

REFERENCES

Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). The variability of human, bold hemodynamic responses. *Neuroimage* 8, 360–369.

Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L., and Sporns, O. (2009). Modeling the impact of lesions in the human brain. *PLoS Comput. Biol.* 5:e1000408. doi: 10.1371/journal.pcbi.1000408

Bakkour, A., Morris, J. C., Wolk, D. A., and Dickerson, B. C. (2013). The effects of aging and Alzheimer's disease on cerebral cortical anatomy: specificity and differential relationships with cognition. *Neuroimage* 76, 332–344. doi: 10.1016/j.neuroimage.2013.02.059

Barnett, L., Barrett, A. B., and Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* 103:238701.

Barnett, L., and Seth, A. K. (2014). The MVGC multivariate granger causality toolbox: a new approach to granger-causal inference. *J. Neurosci. Methods* 223, 50–68. doi: 10.1016/j.jneumeth.2013.10.018

Bressler, S. L., and Seth, A. K. (2011). Wiener–granger causality: a well established methodology. *Neuroimage* 58, 323–329. doi: 10.1016/j.neuroimage.2010.02.059

Chen, C., Tyler, C. W., and Baseler, H. A. (2003). Statistical properties of bold magnetic resonance activity in the human brain. *Neuroimage* 20, 1096–1109.

Chen, Y., Bressler, S. L., and Ding, M. (2006). Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J. Neurosci. Methods* 150, 228–237. doi: 10.1016/j.jneumeth.2005.06.011

Chivukula, A. S., Li, J., and Liu, W. (2018). "Discovering Granger-causal features from deep learning networks," in *AI 2018: Advances in Artificial Intelligence*, Vol. 11320, eds T. Mitrovic, B. Xue, and X. Li (Cham: Springer), 692–705.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signal. Syst.* 2, 303–314. doi: 10.1007/bf02551274

Dauwels, J., Vialatte, F., Musha, T., and Cichocki, A. (2010). A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *Neuroimage* 49, 668–693. doi: 10.1016/j.neuroimage.2009.06.056

Dauwels, J., Vialatte, F., Rutkowski, T., and Cichocki, A. (2008). Measuring neural synchrony by message passing. *Adv. Neural Inform. Process. Syst.* 20, 361–368.

Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G. L., Hagmann, P., and Corbetta, M. (2013). Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *J. Neurosci.* 33, 11239–11252. doi: 10.1523/JNEUROSCI.1091-13.2013

Delalleau, O. and Bengio, Y. (2011). Shallow vs. deep sum-product networks. *Adv. Neural Inf. Process. Syst.* 24, 666–674.

Duggento, A., Guerrisi, M., and Toschi, N. (2019). Echo state network models for nonlinear Granger causality. *bioRxiv*[Preprint] bioRxiv: 651679,

Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302.

Friston, K., Moran, R., and Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* 23, 172–178.

Friston, K. J., Frith, C. D., Liddle, P. F., and Sj Franckowiak, R. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 13, 5–14. doi: 10.1038/jcbfm.1993.4

Gămănuț, R., Kennedy, H., Toroczkai, Z., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., et al. (2018). The mouse cortical connectome, characterized by an ultra-dense cortical graph, maintains specificity by distinct connectivity profiles. *Neuron* 97, 698–715.e10. doi: 10.1016/j.neuron.2017.12.037

Gao, Q., Duan, X., and Chen, H. (2011). Evaluation of effective connectivity of motor areas during motor imagery andexecution using conditional Granger causality. *Neuroimage* 54, 1280–1288. doi: 10.1016/j.neuroimage.2010.08.071

Gerstein, G. L., Bedenbaugh, P., and Aertsen, A. M. H. J. (1989). Neuronal assemblies. *IEEE Trans. Biomed. Eng.* 36, 4–14.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* 77, 304–313. doi: 10.1080/01621459.1982.10477803

Glorot, X., Bordes, A., and Bengio, Y. (2011). "Deep sparse rectifier neural networks," *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, Vol. 15 (Montreal, QC: Department of IRO University of Montreal). 315–323.

Goebel, R., Roebroeck, A., Kim, D. S., and Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging* 21, 1251–1261. doi: 10.1016/j.mri.2003.08.026

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.

Goodfellow, I. J., Bulatov, Y., Ibarz, J., Arnoud, S., and Shet, V. (2014). Multi-digit number recognition from street viewimagery using deep convolutional neural networks. *arXiv*[Preprint] arXiv: 1312.6082.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791

Hafkemeijer, A., Möller, C., Dopper, E. G., Jiskoot, L. C., Schouten, T. M., van Swieten, J. C., et al. (2015). Resting state functional connectivity differences between behavioral variant frontotemporal dementia and Alzheimer's disease. *Front. Hum. Neurosci.* 9:474. doi: 10.3389/fnhum.2015.00474

Hahn, G., Skeide, M. A., Mantini, D., Ganzetti, M., Destexhe, A., Friederici, A. D., et al. (2019). A new computational approach to estimate whole-brain effective connectivity from functional and structural MRI, applied to language development. *Sci. Rep.* 9:8479. doi: 10.1038/s41598-019-44909-6

Harasty, J. A., Halliday, G. M., Kril, J. J., and Code, C. (1999). Specific temporoparietal gyral atrophy reflects thepattern of language dissolution in Alzheimer's disease. *Brain* 122, 675–686. doi: 10.1093/brain/122.4.675

Honey, C. J., Kotter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl Acad. Sci. U. S. A.* 104, 10240–10245. doi: 10.1073/pnas.0701519104

Hornik, K. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw.* 2, 359–366.

Jacobs, H. I. L., Hopkins, D. A., Mayrhofer, H. C., Bruner, E., van Leeuwen, F. W., Raaijmakers, W., et al. (2018). The cerebellum in Alzheimer's disease: evaluating its role in cognitive decline. *Brain* 141, 37–47. doi: 10.1093/brain/awx194

Kingma, D. P., and Ba, J. (2014). Adam: a Method for Stochastic Optimization. *arXiv*[Preprint] arXiv: 1412.6980,

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* 56, 455–475. doi: 10.1016/j.neuroimage.2010.07.034

Krueger, F., Landgraf, S., Meer, E., Deshpande, G., and Hu, X. (2011). Effective connectivity of the multiplication network: a functional MRI and multivariate granger causality mapping study. *Hum. Brain Mapp.* 32, 1419–1431. doi: 10.1002/hbm.21119

Li, X., Morgan, P. S., Ashburner, J., Smith, J., and Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J. Neurosci. Methods* 264, 47–56. doi: 10.1016/j.jneumeth.2016.03.001

Liang, S., and Srikant, R. (2016). Why deep neural networks for function approximation? *arXiv*[Preprint] arXiv: 1610.04161,

Marinazzo, D., Liao, W., Chen, H., and Stramaglia, S. (2011). Nonlinear connectivity by Granger causality. *Neuroimage* 58, 330–338. doi: 10.1016/j.neuroimage.2010.01.099

Marinazzo, D., Pellicoro, M., and Stramaglia, S. (2012). Causal information approach to partial conditioning in multivariate data sets. *Comp. Math. Methods. Med.* 2012:303601. doi: 10.1155/2012/303601

Matsumoto, M., and Nishimura, T. (1998). A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. Model. Comput. Simul.* 8, 3–30.

Melozzi, F., Woodman, M. M., Jirsa, V. K., and Bernard, C. (2017). The virtual mouse brain: a computational neuroinformatics platform to study whole mouse brain dynamics. *eNeuro* 4:ENEURO.0111-17.2017., doi: 10.1523/ENEURO.0111-17.2017

Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi: 10.1007/s00429-010-0262-0

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., and Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage* 11, 735–759. doi: 10.1006/nimg.2000.0568

Montalto, A., Faes, L., and Marinazzo, D. (2014). MuTE: a MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy. *PLoS One* 9:e109462. doi: 10.1371/journal.pone.0109462

Montalto, A., Stramaglia, S., Faes, L., Tessitore, G., Prevete, R., and Marinazzo, D. (2015). Neural networks with non-uniform embedding and explicit validation phase to assess granger causality. *Neural Netw.* 71, 159–171. doi: 10.1016/j.neunet.2015.08.003

Mosconi, L., Pupi, A., and De Leon, M. J. (2009). Brain glucose hypometabolism and oxidative stress in preclinical Alzheimer's disease. *Ann. N. Y. Acad. Sci.* 1147, 180–195.

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Haifa, 807–814.

Oh, S., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., et al. (2014). A mesoscale connectome of the mouse brain. *Nature* 508, 207–214.

Pascoal, T. A., Mathotaarachchi, S., Kang, M. S., Mohaddes, S., Shin, M., and Park, A. Y. (2019). Aβ-induced vulnerability propagates via the brain's default mode network. *Nat. Commun.* 10:2353. doi: 10.1038/s41467-019-10217-w

Peraza, R. P., Kaiser, M., Firbank, M., Graziadio, S., Bonanni, L., and Onofrj, M. (2014). fMRI resting state networks and their association with cognitive fluctuations in dementia with Lewy bodies. *Neuroimage* 4, 558–565. doi: 10.1016/j.nicl.2014.03.013

Prando, G., Zorzi, M., Bertoldo, A., Corbetta, M., Zorzi, M., and Chiuso, A. (2020). Sparse DCM for whole-brain effective connectivity from resting-state fMRI data. *Neuroimage* 208:116367. doi: 10.1016/j.neuroimage.2019.116367

Ryali, S., Chen, T., Supekar, K., and Menon, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage* 59, 3852–3861. doi: 10.1016/j.neuroimage.2011.11.054

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85:461. doi: 10.1103/physrevlett.85.461

Schwab, S., Afyouni, S., Chen, Y., Han, Z., Guo, Q., and Dierks, T. (2020). Functional connectivity alterations of the temporal lobe and hippocampus in semantic dementia and Alzheimer's disease. *J. Alzheimers Dis.* 76, 1461–1475.

Shannon, C. (1948). Amathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.

Shimizu, S., Hoyer, P. O., Hyvarinen, A., and Kerminen, A. (2006). A Linear Non-Gaussian Acyclic Model for Causal Discovery. *J. Mach. Learn. Res.* 7, 2003–2030.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for FMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Stokes, P. A., and Purdon, P. L. (2017). A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proc. Natl. Acad. Sci. U. S. A.* 114, E7063–E7072.

Utevsky, A. V., Smith, D. V., and Huettel, S. A. (2014). Precuneus is a functional core of the default-mode network. *J. Neurosci.* 34, 932–940. doi: 10.1523/jneurosci.4227-13.2014

Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., and Friston, K. (2011). Effective connectivity: influence, causality and biophysical modeling. *Neuroimage* 58, 339–361. doi: 10.1016/j.neuroimage.2011.03.058

Wang, Y., Lin, K., Qi, Y., Lian, Q., Feng, S., Wu, Z., et al. (2018). Estimating brain connectivity with varying-length time lags using a recurrent neural network. *IEEE Trans. Biomed. Eng* 65, 1953–1963. doi: 10.1109/TBME.2018.2842769

Wen, X., Yao, L., Liu, Y., and Ding, M. (2012). Causal interactions in attention networks predict behavioral performance. *J. Neurosci.* 32, 1284–1292. doi: 10.1523/jneurosci.2817-11.2012

Whitfield-Gabrieli, S., and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* 2, 125–141. doi: 10.1089/brain.2012.0073

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chem. Int. Lab. Syst.* 58, 109–130. doi: 10.1016/s0169-7439(01)00155-1

Wong, K. F., and Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* 26, 1314–1328. doi: 10.1523/jneurosci.3733-05.2006

Zhou, Z., Chen, Y., Ding, M., Wright, P., Lu, Z., and Liu, Y. (2009). Analyzing brain networks with PCA and conditional granger causality. *Hum. Brain Mapp.* 30, 2197–2206. doi: 10.1002/hbm.20661

Zott, B., Busche, M. A., Sperling, R. A., and Konnerth, A. (2018). What happens with the circuit in Alzheimer's disease in mice and humans?". *Annu. Rev. Neurosci.* 41, 277–297. doi: 10.1146/annurev-neuro-080317-061725

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* 67(Pt 2), 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 24 November 2021
doi: 10.3389/fnins.2021.797378

Noise Robustness Low-Rank Learning Algorithm for Electroencephalogram Signal Classification

*Ming Gao[1], Runmin Liu[2] and Jie Mao[2]**

[1] College of Sports Science and Technology, Wuhan Sports University, Wuhan, China, [2] College of Sports Engineering and Information Technology, Wuhan Sports University, Wuhan, China

Electroencephalogram (EEG) is often used in clinical epilepsy treatment to monitor electrical signal changes in the brain of patients with epilepsy. With the development of signal processing and artificial intelligence technology, artificial intelligence classification method plays an important role in the automatic recognition of epilepsy EEG signals. However, traditional classifiers are easily affected by impurities and noise in epileptic EEG signals. To solve this problem, this paper develops a noise robustness low-rank learning (NRLRL) algorithm for EEG signal classification. NRLRL establishes a low-rank subspace to connect the original data space and label space. Making full use of supervision information, it considers the local information preservation of samples to ensure the low-rank representation of within-class compactness and between-classes dispersion. The asymmetric least squares support vector machine (aLS-SVM) is embedded into the objective function of NRLRL. The aLS-SVM finds the maximum quantile distance between the two classes of samples based on the pinball loss function, which further improves the noise robustness of the model. Several classification experiments with different noise intensity are designed on the Bonn data set, and the experiment results verify the effectiveness of the NRLRL algorithm.

Keywords: electroencephalogram, epilepsy, noise robustness, low-rank learning, pinball loss function

OPEN ACCESS

Edited by:
Kaijian Xia,
Changshu No.1 People's Hospital, China

Reviewed by:
Yufeng Yao,
Changshu Institute of Technology, China
Tongguang Ni,
Changzhou University, China

****Correspondence:***
Jie Mao
maojie@whsu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *18 October 2021*
Accepted: *05 November 2021*
Published: *24 November 2021*

Citation:
Gao M, Liu R and Mao J (2021) Noise Robustness Low-Rank Learning Algorithm for Electroencephalogram Signal Classification. Front. Neurosci. 15:797378. doi: 10.3389/fnins.2021.797378

INTRODUCTION

Brain computer interface (BCI) is a system that collects the signals from the brain to communicate with computers or other devices (Gummadavelli et al., 2018; Jiang et al., 2020). As an efficient way for the human brain to directly communicate with peripheral devices, the BCI does not need to rely on the peripheral nervous system and muscles. Electroencephalogram (EEG) signals, as a biomarker, play an important role in BCI. EEG is often used in clinical diagnosis to determine the presence and type of epilepsy (Fahimi et al., 2019; Jiang et al., 2019). The epileptic seizure process has several different periods: interictal, pre-seizure, and seizure. The waveform, frequency, and signal characteristics of different stages are different in EEG. Based on the analysis of the characteristics of epilepsy EEG, many studies can generally be divided into two directions: epilepsy detection and epilepsy prediction (Jiang et al., 2017; Gu et al., 2021). The epilepsy detection algorithm uses signal processing, machine learning, and deep learning to extract signal features, and distinguishes the EEG signals between the interictal period and the seizure period. The epilepsy prediction algorithm distinguishes the EEG signals in the pre-seizure period and the seizure period.

The prediction task is more difficult than the detection task. First of all, there is no uniform definition of epilepsy prediction and internal standards in the industry. Secondly, compared with the EEG signal in the seizure period, the signal pattern of the EEG signals in the pre-seizure period and the EEG signals in the intermittent period are more similar, so the algorithm is required to be more robust.

Both epilepsy detection and epilepsy prediction are essentially classification tasks in machine learning (Ni et al., 2020). Several studies focus on the classification of EEG information, which involve both epilepsy detection tasks and epilepsy prediction tasks. Zhou et al. (2018) represented the epilepsy EEG signal into a two-dimensional image, and then they constructed a convolutional neural network to automatically learn the transformed image. This method borrows the idea of image processing to analyze the EEG signals and broaden the method of signal processing. Birjandtalab et al. (2017) performed non-linear dimensionality reduction on EEG signals after extracting time-frequency features. This feature processing method can reflect the non-linear relationship of the data in the process of low-dimensional mapping. Ramakrishnan and Muthanantha (2018) computed the approximate entropy value, the maximum Lyapunov component, and the correlation coefficient dimension on different sub-bands of epilepsy EEG signals, and introduced fuzzy rules to fuzzify the features. The authors believe that fuzzy rules are the natural choice of using human professional knowledge to build machine learning systems, which is closely related to people's way of thinking. Wang et al. (2017) explored multiple bands of the EEG signal by considering the maximum and standard deviation characteristics of each band. Then they constructed the feature vector of the EEG and used a one-to-one self-organization strategy to create a high high-precision epilepsy detection system. Sun et al. (2019) intercepted and analyzed the pre-seizure data, and they used recurrent autoencoders on multivariate signals to extracted EEG features. Liu et al. (2019) transformed the EEG signal into spectral data through the combination of dimensionality reduction and short-time Fourier transform. Then, the authors constructed a shallow convolutional neural network (CNN) network to automatically learn data features. Yu et al. (2020) used the local mean decomposition (LMD) method to obtain the feature matrix of the EEG signals, and they used a CNN model to implement feature extraction and combined Bayesian linear discriminant analysis to obtain the prediction result.

In supervised learning, the support vector machine (SVM) represented by least squares regression (LSR) is a simple and effective method. The core idea of LSR is to learn the non-linear projection from the original data to the feature space, and the obtained projection vector of the original data is also used as the data representation in the label space. For example, discriminative LSR method include multiclass classification (Xiang et al., 2012), groupwise retargeted LSR method (Ling and Geng, 2019), regularized label relaxation linear regression (Fang et al., 2018), double relaxed regression for classification (Han et al., 2019), and so on. For epilepsy data, the scalp EEG data will have more impurity signals and noise signals. Moreover, dimensional explosion and information redundancy problems are common in EEG signals. Learning a discriminatively compact data representation is a very critical problem in pattern recognition. At present, there are many methods based on subspace learning and least squares classifier to learn good classifiers. For example, to combine projection learning with the task of exploring label information, Meng et al. (2020) proposed a constrained discriminative projection learning for joint optimization of subspace learning and classification problems, which used low-rank constraints to learn robust subspaces to connect the original visual features and target output. Subspace learning essentially tries to find a suitable low-dimensional space in which the discriminative representation of the original features is preserved as much as possible. In recent years, low-rank learning has achieved relatively good results in matrix analysis, data recovery, and data denoising. At the same time, low-rank representation is an effective means to describe the structure of high-dimensional data, and it is a generalized form of sparsity in matrix space. That is, low-rank representation can describe the low-dimensional subspace structure of high-dimensional data, thus its component in the subspace becomes the most important factor in characterizing the data. In addition, low-rank representation effectively introduces low-rank constraints into the data matrix, which can help to construct discriminative feature subspaces and eliminate outliers. Inspired by this idea, the noise robustness low-rank learning (NRLRL) algorithm is proposed for EEG signal classification. NRLRL learns a low-rank subspace that connects the original data space and the label space. It fully considers the correlation information and local structure of samples, and it guarantees the minimum rank of the coefficient matrix constructed of data under its self-expression. By integrating the multi-class asymmetric least squares SVM classifier with low-rank representation, NRLRL is insensitive to noise and outliers. The experiments performed on noisy EEG signals are shown that our algorithm is noise robust. NRLRL has several advantages as follows: (1) since the low-rank representation follows the minimum rank criterion, NRLRL is robust when reconstructing the original data with noise and outliers. (2) By full use of supervised information and pinball loss function, an asymmetric least square SVM is jointly learned into our objection function, so that NRLRL explores a robust classifier in the framework of low-rank learning. (3) Local constraints based on low-rank representations are used based on supervision information. The criteria of low-rank representations for minimum within-class and maximum between-classes are adopted to capture the discriminative structure of the data.

BACKGROUND

Low-Rank Representation

Give a set of data samples $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, each sample $\mathbf{x}_i \in \mathbf{R}^d$in $\mathbf{X}$ can be represented as a linear combination of atoms from a dictionary $\mathbf{A}$:

$$\mathbf{X} = \mathbf{AC}, \tag{1}$$

where $\mathbf{C} = [\mathbf{c}_1, ..., \mathbf{c}_n] \in \mathbf{R}^{n\times m}$is the coefficient matrix rank representation.

As a common practice in low-rank learning, the dictionary $\mathbf{A}$ is set to $\mathbf{X}$, i.e.,

$$\mathbf{X} = \mathbf{XC}, \tag{2}$$

Eq. (2) uses the data set itself to represent the data, which is called the self-expression of the data. Each data sample in data set $\mathbf{X}$ can be represented by:

$$\mathbf{x}_i = \sum_{j\neq i} C_{ij}\mathbf{x}_j, \tag{3}$$

By minimizing the rank of the coefficient matrix $\mathbf{C}$, Eq. (1) can be written by:

$$\begin{aligned} &\min_{\mathbf{C}} \ rank(\mathbf{C}), \\ &s.t. \ \mathbf{X} = \mathbf{XC}. \end{aligned} \tag{4}$$

where$rank(\mathbf{C})$ is the rank function of $\mathbf{C}$.

Considering the existence of noise and outliers in the data sample, the structure of the original data $\mathbf{X}$ is taken as two parts: one is the linear combination of the dictionary $\mathbf{X}$ and a low-rank coefficient matrix $\mathbf{C}$, and the second part is noise (error) matrix, i.e.,

$$\mathbf{X} = \mathbf{XC} + \mathbf{E}, \tag{5}$$

Then the low-rating representation can be defined as follows:

$$\begin{aligned} &\min_{C,E} \ rank(\mathbf{C}) + \mu \ ||\mathbf{E}||_{2,0}, \\ &s.t. \ \mathbf{X} = \mathbf{XC} + \mathbf{E}, \end{aligned} \tag{6}$$

where $||\cdot||_{2,0}$means the $\ell_{2,0}$-norm operator. μ is the trade-off parameter. In Eq. (6), the term $||E||_{2,0}$ encourages the sparseness of the error components.

The low-rank optimization problem of Eq. (6) is a non-convex NP-hard problem. To find its unique optimal solution, it is necessary to perform convex relaxation of Eq. (6). The kernel norm is the best convex approximation of the rank function on the unit sphere in the matrix spectral norm (Candès and Recht, 2009). Therefore, the convex kernel norm can be used to approximate the non-convex rank function, and the $\ell_{2,0}$ norm can be relaxed to its $\ell_{2,1}$norm (Raghunandan et al., 2010). Then Eq. (6) can be written as the following convex optimization problem

$$\begin{aligned} &\min_{\mathbf{C,E}} \ ||\mathbf{C}||_* + \mu \ ||\mathbf{E}||_{2,1}, \\ &s.t. \ \mathbf{X} = \mathbf{XC} + \mathbf{E}, \end{aligned} \tag{7}$$

where $||\cdot||_*$ means the nuclear norm operator, and$||\cdot||_{2,1}$means the $\ell_{2,1}$-norm operator.

Asymmetric Least Squares Support Vector Machine

The loss function in the least squares SVM (LS-SVM) pays attention to both the correctly classified and incorrectly classified samples. It minimizes the squared error of the classifier as follows,

$$l(\mathbf{x}) = (1 - \mathbf{w}^T\mathbf{x}–b)^2. \tag{8}$$

In fact, the above loss function is noise sensitive, especially the noises around the separation hyperplane. Many extensions of least squares loss function have been proposed to solve this problem, such as iteratively reweighted least square (Leski, 2015) and asymmetric square function (Leski, 2015), asymmetric squared loss (Huang et al., 2014). Using the statistical property to lower quantile value, the asymmetric squared loss is defined as:

$$L_p^{aLS}(\mathbf{x}) = \begin{cases} p(1 - \mathbf{w}^T\mathbf{x} + b)^2, \ if \ 1 - \mathbf{w}^T\mathbf{x} + b \geq 0 \\ (1-p)(1 - \mathbf{w}^T\mathbf{x} + b)^2, \ if \ 1 - \mathbf{w}^T\mathbf{x} + b < 0 \end{cases} \tag{9}$$

where $\mathbf{w}$ and b are the hyperplane parameter and bias parameter of SVM classifier, respectively. p is the lower quantile value parameter.

The aLS-SVM uses the expectile distance and maximizes the expectile distance between different classes. The aLS-SVM has the following optimization problem:

$$\min_{w,b} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{\alpha}{2}\sum_{i=1}^{n} L_p^{aLS}(1 - y_i(\mathbf{w}^T\mathbf{x}_i + b))^2. \tag{10}$$

where α is the regularization parameter. This optimization problem can be solved by quadratic programming method.

NOISE ROBUSTNESS LOW-RANK LEARNING ALGORITHM

The Noise Robustness Low-Rank Learning Model

Given a set of data points $\mathbf{X} = [\mathbf{x}_1, ..., x_n]$ and their labels $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_n]$ are distributed in K classes. $\mathbf{y}_k = [y_{1,k}, y_{2,k}, ..., y_{n,k}]$ is the class label vector of n training samples associated with the k-th class. Considering the influence of noise or outliers, the main goal of our algorithm is to find the lowest rank representation $\mathbf{C}$ and the best classifier based on $\mathbf{C}$.

First, to increase discrimination capability, the local preservation with label embedding is incorporated into the learning process. Different from the traditional local preservation term in low-rank learning, the label information is embedded into the k-nearest neighborhood relationships. For sample $\mathbf{x}_i$, its low-rank is expressed as $\mathbf{c}_i$. Without considering the label information of the sample, if $\mathbf{x}_j$ is in the k-nearest neighbor of $\mathbf{x}_i$, their corresponding low-rank representations $\mathbf{c}_j$ and $\mathbf{c}_i$ should be closer to each other. Obviously, this strategy is not suitable for classification tasks. Based on the basic classification principles of within-class compactness and between-classes separation, the label information is introduced into the k-nearest neighborhood relationships. The within-class matrix $\mathbf{B}_{within}$ and between-classes matrix $\mathbf{B}_{between}$ are accordingly defined, and their elements can be defined as,

$$B_{within,ij} = \begin{cases} \exp(-\left|\left|\mathbf{x}_i - \mathbf{x}_j\right|\right|^2/t), \ if \ \mathbf{x}_i \in N(\mathbf{x}_j) \ or \ \mathbf{x}_j \in N(\mathbf{x}_i), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad y_i = y_j, \\ 0, \qquad\qquad\qquad otherwise \end{cases} \tag{11}$$

$$B_{between,ij} = \begin{cases} -\exp(-\left|\left|\mathbf{x}_i - \mathbf{x}_j\right|\right|^2/t), & if\ \mathbf{x}_i \in N(\mathbf{x}_j)\ or \\ & \mathbf{x}_j \in N(\mathbf{x}_i),\ y_i \neq y_j, \\ 0, & otherwise \end{cases} \quad (12)$$

where $N(\mathbf{x}_j)$ returns the k-nearest neighbors of $\mathbf{x}_j$.

In the NRLRL, the original data is projected into a low-dimensional subspace by low-rank representation. NRLRL shows the similarity within the class and the difference between classes of the data. To achieve this goal, the label embedded local preservation term is defined as,

$$\begin{aligned} \min_{\mathbf{C}} &= \tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\left|\mathbf{c}_i - \mathbf{c}_j\right|\right|_2^2 (B_{within,ij} - B_{between,ij}) \\ &= Tr(\mathbf{C}\mathbf{B}_{within}\mathbf{C}^T) - Tr(\mathbf{C}\mathbf{B}_{between}\mathbf{C}^T) \\ &= Tr(\mathbf{C}\mathbf{L}\mathbf{C}^T) \end{aligned} \quad (13)$$

where $\mathbf{L} = \mathbf{L}_{between} - \mathbf{L}_{within}$, $\mathbf{L}_{between} = \mathbf{B}_{between} - \mathbf{B}_{between}$, $\mathbf{L}_{within} = \mathbf{B}_{within} - \mathbf{B}_{within}$. $\mathbf{B}_{between}$ and $\mathbf{B}_{within}$ are diagonal matrices, and their elements are $\mathbf{B}_{between,ii} = \sum_{j=1}^{n}\mathbf{B}_{between,ij}$, $\mathbf{B}_{within,ii} = \sum_{j=1}^{n}\mathbf{B}_{within,ij}$, separately. $Tr(\cdot)$ is the trace operator.

The label embedded local preservation term is to ensure that if the nearest neighbors $\mathbf{x}_i$ and $\mathbf{x}_j$ are from the same class, their low-rank codes $\mathbf{c}_i$ and $\mathbf{c}_j$ are also close to each other. At the same time, if the nearest neighbors $\mathbf{x}_i$ and $\mathbf{x}_j$ are from different classes, their low-rank codes $\mathbf{c}_i$ and $\mathbf{c}_j$ are separated as much as possible. In such a low-rank learning stage, the ideal non-linear local structure of EEG data is preserved.

To promote the discriminative ability of low-rank representation vectors, a multi-class aLS-SVM classification term is embedded into the NRLRL algorithm. The multi-class aLS-SVM classification term $\mathbf{L}(C)$includes two parts

$$\begin{aligned} &L(\mathbf{C},\mathbf{W},\mathbf{b}) = \textstyle\sum_{k=1}^{K}(f(\mathbf{C},\mathbf{y}_k,\mathbf{w}_k,b_k) + \alpha\,||\mathbf{w}_k||_2^2), \\ &f(\mathbf{C},\mathbf{y}_k,\mathbf{w}_k,b_k) = \textstyle\sum_{i=1}^{n} l(\mathbf{c}_i, y_{k,i}, \mathbf{w}_k, b_k), \end{aligned} \quad (14)$$

where $l(\mathbf{c}_i, y_{k,i}, \mathbf{w}_k, b_k)$ is the loss function associated with the k-th aLS-SVM.

The squared pinball loss in NRLRL can be written as

$$l(\mu) = \begin{cases} p\mu^2, & \mu \geq 0, \\ (1-p)\mu^2, & \mu < 0, \end{cases} \quad (15)$$

where $\mu_{k,i} = y_{k,i}(\mathbf{w}_k^T\mathbf{c}_i + b_k) - 1$.

Bedding the local preservation with label embedding term and multi-class aLS-SVM classification term into the Eq.(7), the objective function of NRLRL can be written as

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{E},\mathbf{W},\mathbf{b}} ||\mathbf{C}||_* + \lambda\,||\mathbf{E}||_{2,1} + \gamma L(\mathbf{C},\mathbf{W},\mathbf{b}) + \eta Tr(\mathbf{C}\mathbf{L}\mathbf{C}^T), \\ &s.t.\ \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, \\ &\quad \mathbf{1}_n^T\mathbf{C} = \mathbf{1}_n^T, \end{aligned} \quad (16)$$

where λ,γ, and η are regularization parameters.

The loss function term is decomposed into the sum of the loss term of each sample, and it can be seen that the contribution of each data sample to the objective function is linearly cumulative. $L(\mathbf{C},\mathbf{W},\mathbf{b})$ includes a class-by-class loss function term $l(\mathbf{c}_i, y_{k,i}, \mathbf{w}_k, b_k)$on the low-rank representation, so that the obtained lowest-rank representations are highly correlated within the class. EEG signals belonging to the same class usually contain the common discriminative features. The lowest-rank representation obtained by Eq. (16) has the characteristics of strong within-class correlation and between-class difference for classification tasks.

To reduce the time costs, the Frobenius norm is used to replace the nuclear norm, the objective function of NRLRL can be re-written as:

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{E},\mathbf{W},\mathbf{b}} ||\mathbf{C}||_F^2 + \lambda\,||\mathbf{E}||_{2,1} + \gamma L(\mathbf{C},\mathbf{W},\mathbf{b}) + \eta Tr(\mathbf{C}\mathbf{L}\mathbf{C}^T), \\ &s.t.\ \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, \\ &\quad \mathbf{1}_n^T\mathbf{C} = \mathbf{1}_n^T, \end{aligned} \quad (17)$$

For simplicity of expression, combining the two terms $||C||_F^2$ and $Tr(C^TLC)$, Eq. (17) can be written as:

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{E},\mathbf{W},\mathbf{b}} \lambda\,||\mathbf{E}||_{2,1} + \gamma L(\mathbf{C},\mathbf{W},\mathbf{b}) + Tr(\mathbf{C}(\eta\mathbf{L} + \mathbf{I})\mathbf{C}^T), \\ &s.t.\ \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, \\ &\quad \mathbf{1}_n^T\mathbf{C} = \mathbf{1}_n^T, \end{aligned} \quad (18)$$

From Eq. (18), we can see that the NRLRL algorithm consists of three sub-problems, namely the parameters of **C**, **E**, and aLS-SVM classifier. These three sub-problems can be solved alternately until the NRLRL algorithm converges. We use the alternating direction multipliers method (ADMM) (Luo et al., 2017) to solve Eq. (18). The augmented Lagrangian function corresponding to Eq. (18) can be written as:

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{E},\mathbf{W},\mathbf{b}} \lambda\,||\mathbf{E}||_{2,1} + \gamma L(\mathbf{C},\mathbf{W},\mathbf{b}) + Tr(\mathbf{C}(\eta\mathbf{L} + \mathbf{I})\mathbf{C}^T) \\ &+Tr[\boldsymbol{\theta}^T(\mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E})] + Tr[\boldsymbol{\delta}^T(\mathbf{1}_n^T\mathbf{C} - \mathbf{1}_n^T)] \\ &+\tfrac{\mu}{2}\left(||\mathbf{X}\text{-}\mathbf{X}\mathbf{C} - \mathbf{E}||_F^2 + \left|\left|\mathbf{1}_n^T\mathbf{C} - \mathbf{1}_n^T\right|\right|_F^2\right) \end{aligned} \quad (19)$$

where θ and δ are Lagrange multipliers, and μ is the penalty parameter.

Optimization of the Objective Function

According to the ADMM algorithm, the parameters in Eq. (19) can be updated alternately, that is, when one parameter is updated, other parameters are fixed until the NRLRL algorithm converges.

(1) Update **C** by fixing **E**, $\mathbf{w}_k$, and b_k. Eq. (19) is converted to the following problem:

$$\begin{aligned} &L(\mathbf{c}_i) = \gamma\left|\left|1 - y_{k,i}(\mathbf{w}_k^T\mathbf{c}_i + b_k)\right|\right|_2^2 + Tr(\mathbf{c}_i(\eta\mathbf{L} + \mathbf{I})\mathbf{c}_i^T) \\ &+Tr[\theta^T(\mathbf{x}_i - \mathbf{x}_i\mathbf{c}_i) + \delta^T\mathbf{1}_n^T\mathbf{c}_i] \\ &+\tfrac{\mu}{2}\left(||\mathbf{x}_i - \mathbf{x}_i\mathbf{c}_i - \mathbf{e}_i||_F^2 + \left|\left|\mathbf{1}_n^T\mathbf{c}_i - \mathbf{1}_n^T\right|\right|_F^2\right). \end{aligned} \quad (20)$$

Let the derivative of Eq. (19) with respect to c_i be zero, the solution of c_i is:

$$\mathbf{c}_i = \left(2(\eta\mathbf{L}+\mathbf{I}) + \mu(\mathbf{x}_i^T\mathbf{x}_i + \mathbf{1}_n\mathbf{1}_n^T) + 2\gamma\sum \mathbf{w}_k\mathbf{w}_k^T\right)^{-1}$$
$$(\mu(\mathbf{x}_i^T\mathbf{x}_i - \mathbf{x}_i^T\mathbf{e}_i + \mathbf{1}_n\mathbf{1}_n^T) + \mathbf{x}_i^T\theta - 1_n\delta + 2\gamma$$
$$\sum \mathbf{w}_k(y_{k,i} - b_k)) \tag{21}$$

(2) Update **E** by fixing **C**, $\mathbf{w}_k$, and b_k. Using the same calculation and reduction strategy by Liu et al. (2013). Eq. (19) is converted to the following problem:

$$\min_{\mathrm{E}} \frac{\lambda}{\mu}||\mathbf{E}||_{2,1} + \frac{1}{2}\left|\left|\mathbf{E} - (\mathbf{X} - \mathbf{XC} + \frac{1}{\mu}\theta)\right|\right|_F^2, \tag{22}$$

Then the solution of **E** can be obtained by:

$$\mathbf{E}(:,i) = \begin{cases} \frac{||\theta_i||-\lambda}{||\theta_i||}\theta_i, & if\ \frac{\lambda}{\mu} < ||\theta_i|| \\ 0, & otherwise \end{cases}, \tag{23}$$

where $\theta_i$is the ith column vector of the matrix θ .

(3) Update $\mathbf{w}_k$ and b_k by fixing **E** and **C**. Eq. (19) is converted to the following problem:

$$\min_{\mathbf{w}_k, b_k} \sum_{k=1}^{K}\left\{\alpha||\mathbf{w}_k||_2^2 + \sum_{i=1}^{n} l(\mathbf{c}_i, y_{i,k}, \mathbf{w}_k, b_k)\right\}, \tag{24}$$

Eq. (24) is a multi-class aLS-SVM problem, and the optimal parameters $\mathbf{w}_k$ and b_k can be obtained by the aLS-SVM algorithm.

The training procedure of NRLRL is summarized in **Algorithm 1**.

ALGORITHM 1 | The training procedure of NRLRL is summarized.

Algorithm NRLRL: noise robustness low-rank learning algorithm
Input: Training samples **X**, label matrix **Y**.
Output: lowest-rank representation matrix **C**, multi-class aLS-SVM classifier.
$t = 1$
While not convergence or t < maximum number of iterations
  Update **C**(t) by solving Eq. (21);
  Update **E**(t) by solving Eq. (23);
  Update **w**(t) and b(t) by solving aLS-SVM algorithm;
  $t = t + 1$
end while

EXPERIMENT

Datasets and Experimental Settings

This study used EEG signals are from Bonn University. The data set consists of five subsets (groups A to E), each of which consists of 100 EEG segments with a single channel duration of 23.6 s and 4,097 samples. The fragments in groups A-B were taken from five healthy subjects, and the fragments in groups {C, D, E} were taken from patients with epilepsy. The groups C and D recorded the signal during the intermittent period of epileptic seizures. The group E recorded the signal during the seizure. The signals of the five groups of EEG data are shown in **Figure 1**. In the experiment, the 4,097 data points were divided into three data blocks to obtain the research samples, that is, a data block is a sample, representing the EEG information in about 8 s. Therefore, the sample size of this paper is 3 × 100 = 300 in each group, and each sample has 1,365 features of sampling points. We design two types of classification tasks on the Bonn dataset. One is the binary classification task: non-epileptic condition (sets {A, B, C, D, E}) and epileptic condition (set E). The other is the three classes of classification task: normal (sets {A, B}), interictal (sets {C, D}), and ictal (set E).

We use the following algorithms as the comparison algorithms: DLSR (Xiang et al., 2012), LC-KSVD (Jiang et al., 2013), SRRS (Li and Fu, 2016), LRSD (Kong et al., 2017), aLS-SVM (Huang et al., 2014), and LRDLSR (Chen et al., 2020). The parameters of these comparison methods are set to their default settings. In NRLRL, the dictionary size is set from {40, …, 320}, three regularization parameters are set from $\{2^{-5}, \ldots, 2^3\}$, the k-nearest neighbor parameter is set from {3, …, 11}, and the pinball loss parameter p is set from {40, …, 360}. We adopt the one-by-one strategy to select the optimal parameters.

Following the method of references (Huang et al., 2014; Gu et al., 2019, 2020), 20 and 50% samples are randomly selected and common Gaussian white noise is added. To test the sensitivity of the classifier to noise intensity, the intensity of Gaussian white noise is divided into three types: the mean value is 0, and the variance is set to 5, 10, and 15% of the sample features, respectively. For example, the *noise* (20%, 10%) indicates that 20% of the samples in the Bonn dataset contain Gaussian white noise, and the variance of the noise is 10% of the sample features.

Classification Result Comparison

First, we perform experiments on the binary classification task. We compare all algorithms in indexes of specificity, sensitivity, and accuracy on the noisy Bonn dataset. The experimental results of the binary classification task are shown in **Tables 1–3**.

From the experimental results, it can be seen that (1) with the increase of noise intensity, the specificity, sensitivity, and accuracy of all algorithms show a decline in varying degrees. It can be seen that the characteristic noise of samples will seriously affect the classification effect of the classifier. Especially the DLSR and LC-KSVD algorithms do not consider the impact of the noise sample interference on the classification surface. As the noise intensity increases, the classification result decreases rapidly. (2) SRRS, LRSD, aLS-SVM, LRDLSR, and NRLRL algorithms are all noise-insensitive classification algorithms, therefore the classification results are significantly better than conventional classification algorithms. The proposed NRLRL algorithm achieves the best results in classification performance. The NRLRL algorithm removes the influence of noise on the sample in the lowest rank representation, and it uses the pinball loss function to obtain a noise-insensitive classification classifier

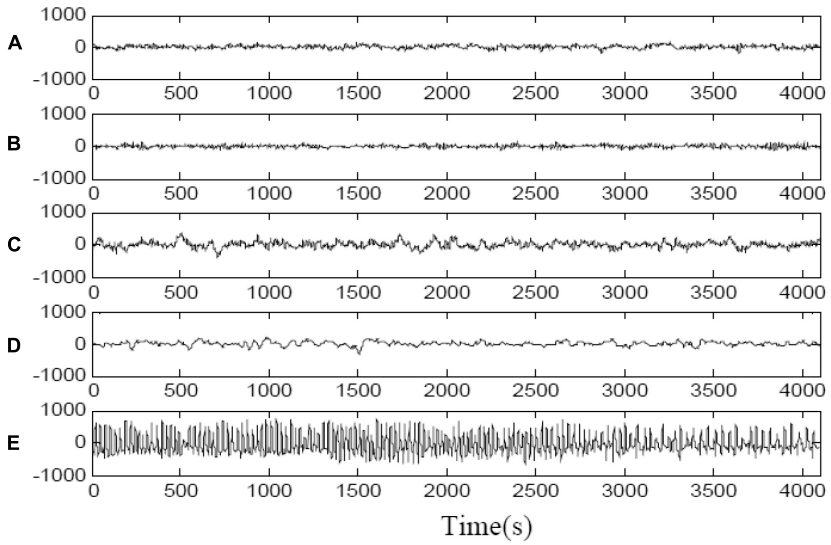


FIGURE 1 | Sample electroencephalogram (EEG) signals in each group in Bonn dataset. **(A)** Epileptic EEG signals measured from healthy people with eyes open. **(B)** Epileptic EEG signals measured from healthy people with eyes closed. **(C)** Epileptic EEG signals obtained in hippocampal formation of the opposite hemisphere of brain during seizure free intervals. **(D)** Epileptic EEG signals obtained from within epileptogenic zone during seizure free intervals. **(E)** Epileptic EEG signals measured during seizure activity.

by maximizing the distance between the two classes of quantile distances. In addition, the NRLRL algorithm can mine the geometric structure of samples in low-dimensional space by low-rank learning, and fully considers the correlation information and subspace structure between samples. Therefore, the within-class similarity and between-class differences of the data are more prominent, which makes the NRLRL algorithm obtain good classification performance in the presence of noise.

Then, we perform experiments on three class classification task, i.e., classification of EEG data in normal, interictal and epileptic periods. Similarly to the above experiment procedure, we compare all algorithms in indexes of specificity, sensitivity, and accuracy on the noisy Bonn dataset. The experimental results of three classification task are shown in **Tables 4**–**6**. From these results, it can be observed that since the complexity of the three class classification is higher than that of the binary class classification, the results of three class classification are lower than those of binary class classification. The proposed NRLRL algorithm achieves the best results of specificity, sensitivity, and accuracy.

Parameter Analysis

Here we discuss the key parameters in the NRLRL algorithm on binary classification task *noise* (20%, 10%) and three class classification *noise* (50%, 10%).

The k-nearest parameter k-nearest is an important parameter in NRLRL. It determines the neighbor relationship between samples. The k-nearest parameter is set from {3, …, 11}. The

TABLE 2 | The sensitivity results of binary classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	92.94	93.49	95.31	95.50	95.74	95.96	**96.65**
noise (20%, 10%)	92.63	93.18	95.24	95.39	95.52	95.91	**96.63**
noise (20%, 15%)	92.08	92.91	95.27	95.25	95.40	95.84	**96.58**
noise (50%, 5%)	90.04	91.63	94.71	95.11	95.28	95.54	**96.52**
noise (50%, 10%)	90.42	91.02	94.52	94.31	94.72	95.31	**96.48**
noise (50%, 15%)	89.95	90.47	94.16	94.39	94.77	95.15	**96.46**

The bold values mean the best values in comparison experiments.

TABLE 1 | The specificity results of binary classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	92.93	93.06	95.08	95.06	95.57	95.50	**96.55**
noise (20%, 10%)	92.61	92.70	95.05	95.21	95.32	95.61	**96.53**
noise (20%, 15%)	92.04	92.87	95.06	95.01	94.80	95.46	**96.51**
noise (50%, 5%)	90.97	91.11	94.61	94.96	95.06	95.30	**96.40**
noise (50%, 10%)	90.40	90.54	94.36	94.29	94.62	95.23	**96.36**
noise (50%, 15%)	89.84	90.32	93.99	94.31	94.49	95.06	**96.33**

The bold values mean the best values in comparison experiments.

TABLE 3 | The accuracy results of binary classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	92.93	93.22	95.24	95.20	95.57	95.69	**96.58**
noise (20%, 10%)	92.61	92.72	95.16	95.33	95.34	95.76	**96.57**
noise (20%, 15%)	92.07	92.88	95.19	95.19	94.99	95.48	**96.54**
noise (50%, 5%)	90.03	91.29	94.62	95.02	95.18	95.33	**96.42**
noise (50%, 10%)	90.41	90.69	94.47	94.28	94.63	95.26	**96.41**
noise (50%, 15%)	89.90	90.46	94.06	94.38	94.57	95.09	**96.39**

The bold values mean the best values in comparison experiments.

TABLE 4 | The specificity results of three class classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	89.41	89.43	92.45	92.78	92.66	93.31	**94.18**
noise (20%, 10%)	89.20	89.15	91.06	92.26	91.65	93.27	**94.11**
noise (20%, 15%)	88.34	88.45	91.72	92.04	91.90	93.08	**93.96**
noise (50%, 5%)	87.59	87.28	92.92	91.77	91.27	92.41	**93.69**
noise (50%, 10%)	86.85	86.87	91.67	91.46	92.19	92.44	**93.62**
noise (50%, 15%)	85.30	85.35	91.41	91.13	91.15	92.83	**93.50**

The bold values mean the best values in comparison experiments.

TABLE 5 | The sensitivity results of three class classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	89.51	89.51	92.60	92.85	92.69	93.39	**94.20**
noise (20%, 10%)	89.22	89.26	92.1	92.31	92.74	93.32	**94.14**
noise (20%, 15%)	88.66	88.49	91.99	92.03	92.08	93.27	**93.99**
noise (50%, 5%)	87.69	87.33	92.53	91.78	91.42	92.55	**93.72**
noise (50%, 10%)	86.88	86.81	91.79	91.51	92.21	92.54	**93.65**
noise (50%, 15%)	86.49	86.42	91.49	91.19	91.28	92.85	**93.53**

The bold values mean the best values in comparison experiments.

TABLE 6 | The accuracy results of three class classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	89.46	89.48	92.48	92.80	92.67	93.32	**94.18**
noise (20%, 10%)	89.20	89.17	92.04	92.28	92.61	93.29	**94.01**
noise (20%, 15%)	88.54	88.48	91.77	92.03	91.92	93.19	**93.95**
noise (50%, 5%)	87.57	87.28	92.47	91.64	91.37	92.49	**93.70**
noise (50%, 10%)	86.86	86.82	91.63	91.43	92.18	92.47	**93.63**
noise (50%, 15%)	86.40	86.39	91.43	91.14	91.22	92.82	**93.52**

The bold values mean the best values in comparison experiments.

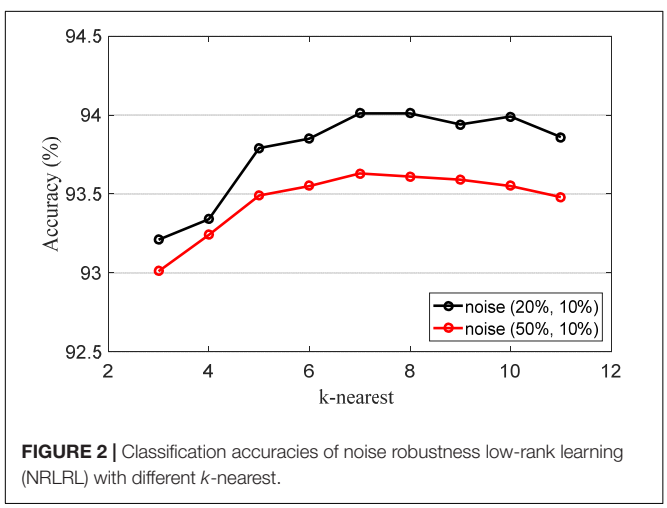


FIGURE 2 | Classification accuracies of noise robustness low-rank learning (NRLRL) with different *k*-nearest.

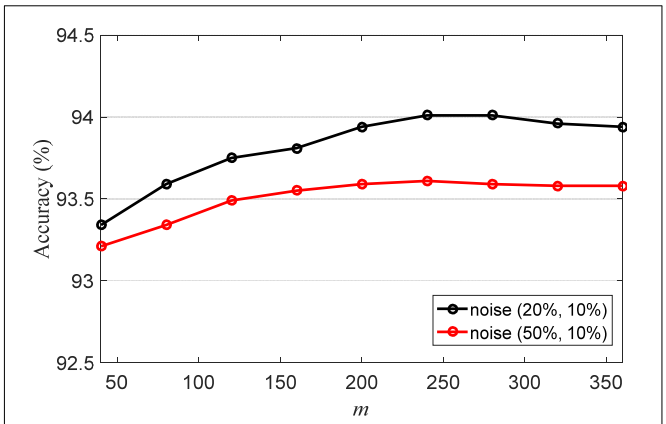


FIGURE 3 | Classification accuracies of noise robustness low-rank learning (NRLRL) with different *m*.

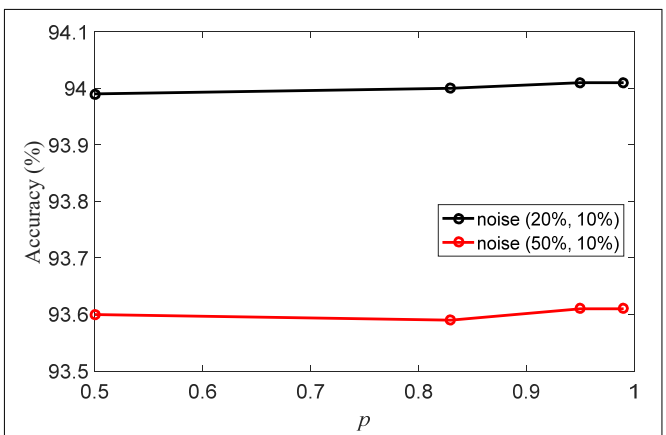


FIGURE 4 | Classification accuracies of noise robustness low-rank learning (NRLRL) with different *p*.

classification accuracies of NRLRL with different *k*-nearest are shown in **Figure 2**. When $k = 7$, the classification accuracy is the highest. The appropriate value of k can reflect the local structural information of the sample to the greatest extent. From the results in **Figure 2**, we can set $k = 7$ in the NRLRL algorithm for noisy Bonn dataset.

Another important parameter is m, which is the size of the matrix **C**. The classification accuracies of NRLRL with different m are shown in **Figure 3**. The parameter m controls the data structure of the low-rank space. When m is too small, the low-rank representation related to the data is not enough to model its structure in the low-rank space. When m is too large, the redundant information will produce errors of low-rank representation. From the results in **Figure 3**, we can set $m = 240$.

The pinball loss parameter p is the important parameter in aLS-SVM classifier in NRLRL. The value range of the pinball loss parameter is {0.5, 0.83, 0.95, 0.99}. The classification accuracies of NRLRL with different p are shown in **Figure 4**. With different values of p, the NRLRL algorithm has achieved high classification accuracy. It shows that the NRLRL algorithm is not sensitive to the p parameter, so the value of p can be fixed to 0.95 in the experiment.

CONCLUSION

In this study, the NRLRL algorithm is proposed for EEG signal classification. Different from noise-insensitive SVM learns the classification hyperplane in the original space or kernel space, NRLRL learns a low-rank subspace as the transformation from the original data space to the label space, to improve the overall classification effect. By introducing the criteria of low-rank representations for minimum within-class and maximum between-classes, the discriminative ability of the model has been greatly improved. The pinball loss function is also helpful to improve the noise insensitivity of the model. The effectiveness of the proposed algorithm is verified on the noisy Bonn EEG dataset. Since our algorithm directly uses the EEG sample point as the input features, we will consider applying various feature extraction methods to the NRLRL algorithm in the next stage. The seizure data is often insufficient in epilepsy detection or prediction tasks. To obtain an effective algorithm model, the down-sampling strategy is often performed to make the class balanced. This strategy will cause the loss of signal data. Therefore, research on appropriate imbalanced data classification methods is the focus of the next stage.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/benfulcher/hctsaTutorial_BonnEEG.

AUTHOR CONTRIBUTIONS

MG conceived and designed the proposed model and wrote the manuscript. RL and JM performed the experiment. All authors read and approved the manuscript.

FUNDING

This work was supported in part by the Scientific Research Program of Education Department of Hubei Province, China (D20184101), the Higher Education Reform Project of Hubei Province, China (201707), the "East Lake Scholar" of Wuhan Sport University Fund, China, and the Hubei Provincial University Specialty subject group construction Special fund, China.

REFERENCES

Birjandtalab, J., Pouyan, M. B., Cogan, D., Nourani, M., and Harvey, J. (2017). Automated seizure detection using limited-channel EEG and non-linear dimension reduction. *Comput. Biol. Med.* 82, 49–58. doi: 10.1016/j.compbiomed.2017.01.011

Candès, E. J., and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* 9:717. doi: 10.1007/s10208-009-9045-5

Chen, Z., Wu, X. J., and Kittler, J. (2020). Low-rank discriminative least squares regression for image classification. *Signal Process.* 173:107485. doi: 10.1016/j.sigpro.2020.107485

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T. S., Ang, K. K., and Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *J. Neural Eng.* 16:026007. doi: 10.1088/1741-2552/aaf3f6

Fang, X., Xu, Y., Li, X., Lai, Z., Wong, W. K., and Fang, B. (2018). Regularized label relaxation linear regression. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 1006–1018. doi: 10.1109/TNNLS.2017.2648880

Gu, X. Q., Ni, T. G., and Fan, Y. Q. (2019). A fast and robust support vector machine with anti-noise convex hull and its application in large-scale ncRNA data classification. *IEEE Access* 7, 134730–134741. doi: 10.1109/ACCESS.2019.2941986

Gu, X. Q., Ni, T. G., Fan, Y. Q., and Wang, W. B. (2020). Scalable kernel convex hull online support vector machine for intelligent network traffic classification. *Ann. Telecommun.* 75, 471–486. doi: 10.1007/s12243-020-00767-2

Gu, X. Q., Zhang, C., and Ni, T. G. (2021). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1679–1687. doi: 10.1109/TCBB.2020.3006699

Gummadavelli, A., Zaveri, H. P., Spencer, D. D., and Gerrard, J. L. (2018). Expanding brain-computer interfaces for controlling epilepsy networks: novel thalamic responsive neurostimulation in refractory epilepsy. *Front. Neurosci.* 12:474. doi: 10.3389/fnins.2018.00474

Han, N., Wu, J., Fang, X. Z., Wong, W. K., Xu, Y., Yang, J., et al. (2019). Double relaxed regression for image classification. *IEEE Trans. Circ. Syst. Video Technol.* 30, 307–319. doi: 10.1109/TCSVT.2018.2890511

Huang, X., Lei, S., and Suykens, J. (2014). Asymmetric least squares support vector machine classifiers. *Computat. Statist. Data Anal.* 70, 395–405. doi: 10.1016/j.csda.2013.09.015

Jiang, Y. Z., Chung, F., and Wang, S. (2019). Recognition of multiclass epileptic EEG signals based on knowledge and label space inductive transfer. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 630–642. doi: 10.1109/TNSRE.2019.2904708

Jiang, Y. Z., Deng, Z. H., Chung, F., Wang, G. J., Qian, P. J., Choi, K., et al. (2017). Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/TFUZZ.2016.2637405

Jiang, Y. Z., Gu, X., Ji, D., Qian, P., Xue, J., Zhang, Y., et al. (2020). Smart diagnosis: a multiple-source transfer TSK fuzzy system for EEG seizure identification. *ACM Trans. Multimed. Comput. Commun. Appl.* 16, 1–21. doi: 10.1145/3340240

Jiang, Z., Lin, Z., and Davis, L. S. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88

Kong, W., Liu, Y., Jiang, B., Dai, G., and Xu, L. (2017). "A new EEG signal processing method based on low-rank and sparse decomposition," in *Proceedings of the Cognitive Systems and Signal Processing. ICCSIP 2016. Communications in Computer and Information Science*, eds F. Sun, H. Liu, and D. Hu (Singapore: Springer), doi: 10.1007/978-981-10-5230-9_54

Leski, J. M. (2015). Fuzzy (c+p)-means clustering and its application to a fuzzy rule-based classifier: toward good generalization and good interpretability. *IEEE Trans. Fuzzy Syst.* 23, 802–812. doi: 10.1109/TFUZZ.2014.2327995

Li, S., and Fu, Y. (2016). Learning robust and discriminative subspace with low-rank constraints. *IEEE Trans. Neural Netw. Learn. Syst.* 27, 2160–2173. doi: 10.1109/TNNLS.2015.2464090

Ling, M. G., and Geng, X. (2019). Soft video parsing by label distribution learning. *Front. Comput. Sci.* 13, 302–317. doi: 10.1007/s11704-018-8015-y

Liu, G., Lin, Z., Yan, S., Yu, Y., and Ma, Y. (2013). Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 171–184. doi: 10.1109/TPAMI.2012.88

Liu, Q., Cai, J., Fan, S. Z., Abbod, M. F., Shieh, J. S., Kung, Y., et al. (2019). Spectrum analysis of EEG signals using CNN to model patient's consciousness level based on anesthesiologists' experience. *IEEE Access* 7, 53731–53742. doi: 10.1109/ACCESS.2019.2912273

Luo, M., Zhang, L., Liu, J., Guo, J., and Zheng, Q. (2017). Distributed extreme learning machine with alternating direction method of multiplier. *Neurocomputing* 261, 164–170. doi: 10.1016/j.neucom.2016.03.112

Meng, M., Lan, M., Yu, J., Wu, J., and Tao, D. (2020). Constrained discriminative projection learning for image classification. *IEEE Trans. Image Process.* 29, 186–198. doi: 10.1109/TIP.2019.2926774

Ni, T. G., Gu, X. Q., and Zhang, C. (2020). An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning. *Front. Neurosci.* 14:837. doi: 10.3389/fnins.2020.00837

Raghunandan, H. K., Andrea, M., and Sewoong, O. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* 11, 2057–2078. doi: 10.1007/s10883-010-9100-1

Ramakrishnan, S., and Muthanantha, M. A. S. (2018). Epileptic seizure detection using fuzzy-rules-based sub-band specific features and layered multi-class SVM. *Pattern Anal. Appl.* 22, 1161–1176. doi: 10.1007/s10044-018-0691-6

Sun, L., Jin, B., Yang, H., Tong, J., Liu, C., and Xiong, H. (2019). Unsupervised EEG feature extraction based on echo state network. *Inform. Sci.* 475, 1–17.

Wang, Y., Li, Z., Feng, L., Zheng, C., and Zhang, W. (2017). Automatic detection of epilepsy and seizure using multiclass sparse extreme learning machine classification. *Comput. Math. Methods Med.* 2017:6849360. doi: 10.1155/2017/6849360

Xiang, S. M., Nie, F. P., Meng, G. F., Pan, C., and Zhang, C. (2012). Discriminative least squares regression for multiclass classification and feature selection. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 1738–1754. doi: 10.1109/TNNLS.2012.2212721

Yu, Z., Nie, W., Zhou, W., Xu, F., Yuan, S., Leng, Y., et al. (2020). Epileptic seizure prediction based on local mean decomposition and deep convolutional neural network. *J. Supercomput.* 76, 3462–3476. doi: 10.1007/s11227-018-2600-6

Zhou, M., Tian, C., Cao, R., Wang, B., Yan, N., Ting, H., et al. (2018). Epileptic seizure detection based on EEG signals and CNN. *Front. Neuroinform.* 12:95. doi: 10.3389/fninf.2018.00095

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 26 November 2021
doi: 10.3389/fnins.2021.782968

Deep Convolutional Neural Network With a Multi-Scale Attention Feature Fusion Module for Segmentation of Multimodal Brain Tumor

Xueqin He[1], Wenjie Xu[1], Jane Yang[2], Jianyao Mao[3], Sifang Chen[3]* and Zhanxiang Wang[4,5]*

[1] School of Informatics, Xiamen University, Xiamen, China, [2] Department of Cognitive Science, University of California, San Diego, San Diego, CA, United States, [3] Department of Neurosurgery, The First Affiliated Hospital of Xiamen University, Xiamen, China, [4] Xiamen Key Laboratory of Brain Center, Department of Neurosurgery, The First Affiliated Hospital of Xiamen University, Xiamen, China, [5] Department of Neuroscience, School of Medicine, Institute of Neurosurgery, Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Yugen Yi,
Jiangxi Normal University, China
Jian Su,
Nanjing University of Information Science and Technology, China

***Correspondence:**
Jane Yang
j7yang@ucsd.edu
Sifang Chen
csfsong143@aliyun.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *25 September 2021*
Accepted: *02 November 2021*
Published: *26 November 2021*

Citation:
He X, Xu W, Yang J, Mao J, Chen S and Wang Z (2021) Deep Convolutional Neural Network With a Multi-Scale Attention Feature Fusion Module for Segmentation of Multimodal Brain Tumor. Front. Neurosci. 15:782968. doi: 10.3389/fnins.2021.782968

As a non-invasive, low-cost medical imaging technology, magnetic resonance imaging (MRI) has become an important tool for brain tumor diagnosis. Many scholars have carried out some related researches on MRI brain tumor segmentation based on deep convolutional neural networks, and have achieved good performance. However, due to the large spatial and structural variability of brain tumors and low image contrast, the segmentation of MRI brain tumors is challenging. Deep convolutional neural networks often lead to the loss of low-level details as the network structure deepens, and they cannot effectively utilize the multi-scale feature information. Therefore, a deep convolutional neural network with a multi-scale attention feature fusion module (MAFF-ResUNet) is proposed to address them. The MAFF-ResUNet consists of a U-Net with residual connections and a MAFF module. The combination of residual connections and skip connections fully retain low-level detailed information and improve the global feature extraction capability of the encoding block. Besides, the MAFF module selectively extracts useful information from the multi-scale hybrid feature map based on the attention mechanism to optimize the features of each layer and makes full use of the complementary feature information of different scales. The experimental results on the BraTs 2019 MRI dataset show that the MAFF-ResUNet can learn the edge structure of brain tumors better and achieve high accuracy.

Keywords: magnetic resonance imaging (MRI), semantic segmentation, convolutional neural network, residual network, attention mechanism, brain tumor

INTRODUCTION

In daily life, the human brain is the controller of all behaviors and the sender of activity instructions. As the main part of the human brain, the cerebrum is the highest part of the central nervous system. Brain health has an important impact on the human body. The brain tumor is one of the most common brain diseases and can be induced at any age. Therefore, the prevention of

brain tumors is a significant part of daily health management. Brain tumors can be divided into glioma, meningioma, pituitary adenoma, schwannoma congenital tumor, and so on, among which glioma accounts for the largest proportion. More than half of gliomas are malignant tumors, among which glioblastoma is the most common malignant tumor of the brain and central nervous system, accounting for about 14.5% of all tumors (Ostrom et al., 2020). According to the World Health Organization (WHO) criteria, gliomas are classified into four grades, and the higher the grade, the more likely the tumor is to be malignant. Among them, grade I and II gliomas are low-grade gliomas (LGG), while grade III and IV gliomas are high-grade gliomas (HGG; Louis et al., 2007), which are malignant tumors. Brain tumors can cause serious damage not only to the brain but also to other parts of the body, such as vision loss, motor problems, sensory problems, and even shock in severe cases. Therefore, early detection of brain tumors and early intervention are the only way to minimize the impact of brain tumors.

In clinical medicine, brain tumor screening and diagnosis mainly involve physical examination, imaging examination, and pathological examination, among which physical examination is a preliminary diagnosis of the patient's condition through a comprehensive physical examination by the doctor. The results are somewhat accidental, and it is impossible to accurately judge the condition. However, the pathological examination requires anesthesia operation to collect samples from patients, which is complicated, costly, and has certain damage to the patient's body. Compared with the previous two methods, medical imaging has the characteristics of objectivity, accuracy, convenience, and low cost. It not only overcomes the inaccuracy and subjectivity of physical examination but also omits the cumbersome collection of biopsy samples in the pathological examination. And it is one of the main methods of auxiliary diagnosis for patients with brain tumors. The medical imaging techniques used to diagnose brain tumors mainly include magnetic resonance imaging (MRI) and computer tomography (CT). MRI images are clearer than CT. Especially for small tumors, the use of CT technology is prone to miss the diagnosis. And for soft tissues, the resolution of CT is much lower than that of MRI. Thus, the results of auxiliary diagnosis and treatment using MRI will be more accurate. In addition, CT imaging requires prior injection of radioactive isotopes into the patient, which can affect the human body to a certain extent. As a non-invasive and low-cost medical imaging technology, MRI has become the first choice for brain tumors diagnosis. In this article, MRI images are utilized as a data carrier to study the segmentation of glioma, which has the greatest risk of malignancies in brain tumors.

The number of brain tumor patients is increasing with the development of society, the accelerated pace of life, and the increase of people's work pressure. Faster and more accurate intervention is the key to reduce the mortality rate of brain tumor-related diseases. During the analysis of the brain images, accurate identification of tumor area is the premise of subsequent qualitative diagnosis. However, large spatial and structural variability and low image contrast are the main problems in brain tumor segmentation.

Traditional brain imaging diagnosis mainly relies on manual analysis by professional doctors, which requires a lot of time and cost. With the huge and increasing amount of medical image data, the speed of manual analysis is far behind the speed of data generation. At the same time, due to the professional knowledge requirements of manual segmentation of brain tumors, the differences and workload of manual segmentation results, machine-participated semi-automatic or fully automatic brain tumor segmentation shows obvious advantages (Gordillo et al., 2013). In early studies, it was mainly aimed at semi-automatic segmentation of brain tumors (Gordillo et al., 2013). The purpose of semi-automatic segmentation research is to minimize human intervention when machines and humans work together to achieve the desired segmentation effect. But it is still affected by differences in human subjective consciousness. The automatic method exploits the model and prior knowledge to achieve independent segmentation.

Segmentation methods for brain tumors can be divided into four categories, which are threshold-based, region-based, classification-based, and model-based methods (Gordillo et al., 2013). Dawngliana et al. (2015) combined initial segmentation of multi-layer threshold with the morphological operation of level set to extract fine images. Since brain tumors are relatively easy to identify compared with other brain tissues, the characteristics of tumor regions can be extracted during the preprocessing stage, so that brain tumors can be segmented using region-based methods. Harati et al. (2011) proposed an improved scale-based fuzzy connectedness algorithm that automatically selects seed points on the scale. The method performed well in low-contrast tumor areas. Region-based methods are greatly affected by image pixel coherence, and noise or intensity changes may lead to holes or excessive segmentation (Gordillo et al., 2013). Another relatively similar idea is based on the prominent characteristics of brain tumors in medical images, which is to segment brain tumors based on tumor contour by feature extraction of brain tumor boundary information (Bauer et al., 2013). Essadike et al. (2018) determined the initial contour by using a tumor filter in analog optics and utilized this initial contour to define the active contour model to determine the tumor boundary. Ma et al. (2018) combined random forest and active contour models to automatically infer glioma structure from multimodal volumetric MR images and proposed a new multiscale patch-driven active contour model to refine the results using sparse representation techniques. In addition, due to the different formation mechanisms and surface features of different brain tumors, many researchers have studied the texture features of different brain tumors, and achieve tumor segmentation through voxel classification or clustering. Among the segmentation methods based on classification, Fuzzy C-means (FCM) is one of the mainstream methods because of its advantages in preserving the original image information. In the early stage, Pham et al. (1997) and Xu et al. (1997) applied the FCM method to MRI segmentation (Latif G. et al., 2021). Subsequently, many variants of standard FCM, such as bias-corrected FCM (BCFCM), enhanced FCM (EFCM), kernelized FCM (KFCM), and spatially constrained KFCM (SKFCM) emerged (Latif G. et al., 2021). However,

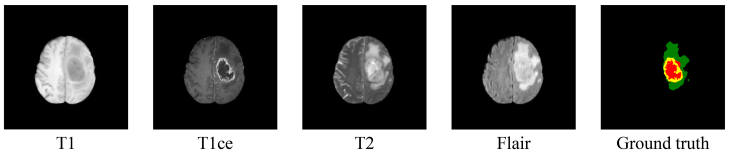


FIGURE 1 | Samples of MRI images in four modalities and their ground truth. In the ground truth image, red, green, and yellow stand for tumor core (TC), whole tumor (WT), and enhance tumor (ET), respectively.

the FCM method is easily disturbed by noise and has a high computational cost. Model-based methods include parametric deformable models, geometric deformable models or level sets, and so on. The above active contour models belong to the parameter deformable model. However, parametric deformable models are difficult to deal with topology changes of contour segmentation and merger naturally, so geometric deformable models or level sets are introduced (Gordillo et al., 2013). Lee et al. (2012) exploited the surface evolution principle of geometric deformation model and level set to achieve medical volume image segmentation and carried out tests on tumor tissues, but the computational efficiency of this method was low.

With the rise of deep learning, researchers began to apply deep networks to the automatic segmentation of brain tumors. Havaei et al. (2017) proposed a brain tumor segmentation model based on a deep neural network, which utilized local features and more global context features to learn the unique features of brain tumor segmentation. For images, a convolutional neural network shows obvious superiority. Pereira et al. (2016) designed a deeper network for glioma segmentation based on a convolutional neural network, using small kernels. Through intensity normalization and data enhancement, the segmentation effect of the network can be improved and the over-fitting can be avoided while the network parameters are minimized. Among them, U-Net (Ronneberger et al., 2015), as the classical model of the convolutional neural network, has outstanding application effect in medical images, so it is also widely used in MRI brain tumor segmentation, and there are many modifications based on U-Net. Latif U. et al. (2021) improved the automatic segmentation process of brain tumors by introducing size variability into the convolutional neural network and proposed a multi-inception-UNET model to improve the scalability of U-Net. Zhang et al. (2020) proposed a new type of densely connection inception convolutional neural network on the basis of U-Net architecture which was applied to medical images, and conducted experiments in tumor segmentation of brain MRI. They added the Inception-Res module and the densely connecting convolutional module to increase the width and depth of the network, and at the same time led to an increase in the number of parameters, which slows down the speed of model training data (Angulakshmi and Deepa, 2021).

In this article, a deep convolutional neural network composed of a U-Net and a multi-scale attention feature fusion module (MAFF) is proposed to achieve automatic segmentation of gliomas in 3D brain MRI images. By using multi-modal MRI data, high-precision segmentation of three tumor types is realized. The main contributions of this work are as follows:

(1) We introduce five residual connections to U-Net, which enhance the feature extraction ability of encoder blocks, the speed of network convergence, and alleviate the gradient vanishing problem caused by the deep network structure.
(2) The proposed MAFF module exploits the attention mechanism to selectively extract feature information of each scale, which can gain a global contextual view. The fusion of useful multi-scale features further improves the accuracy of brain tumor segmentation.
(3) MAFF-ResUNet performs well on the public BraTs 2019 MRI dataset and has certain competitiveness in the field of brain tumor segmentation.

MATERIALS AND METHODS

Dataset

We performed our experiments on the MICCAI BraTs 2019 MRI dataset (Menze et al., 2015; Bakas et al., 2017a,b, 2018). The BraTs 2019 dataset is a collection of MRI data from glioma patients. There are two types of brain tumors in the dataset: high-grade glioma (HGG) and low-grade glioma (LGG). The dataset consists of 256 HGG cases and 76 LGG cases. Each case includes four 3D MRI modalities (T1, T1ce, T2, and Flair) as can be shown in **Figure 1**. And the size of each 3D MRI image is $155 \times 240 \times 240$. The ground truth of each image is labeled manually by the expert. There are four types of labels: background (labeled 0), necrosis and non-enhancing tumor (labeled 1), edema (labeled 2), and enhancing tumor (labeled 4). The task is to segment three nested subregions generated by the three labels (1, 2, and 4), named enhancing tumor (ET, the region of label 4), whole tumor (WT, the region consists of label 1, 2, and 4) and tumor core (TC, the region of label 1 and 4).

Preprocessing

In this work, 3D MRI images from 335 cases in the BraTs 2019 dataset are sliced into multiple 2D images, and slices without tumors are excluded. We use 80% of the generated slices for training, 10% for validation, and 10% for testing. Compared with single modal data, multi-modal data provide more characteristic information for tumor segmentation. To make effective use of multi-modal image information, we concatenate MRI 2D images of four modes in the same dimension as model input.

It is necessary to preprocess the input image before training. First, we remove the top 1% and bottom 1% intensities as Havaei did (Havaei et al., 2017). Then, since images of

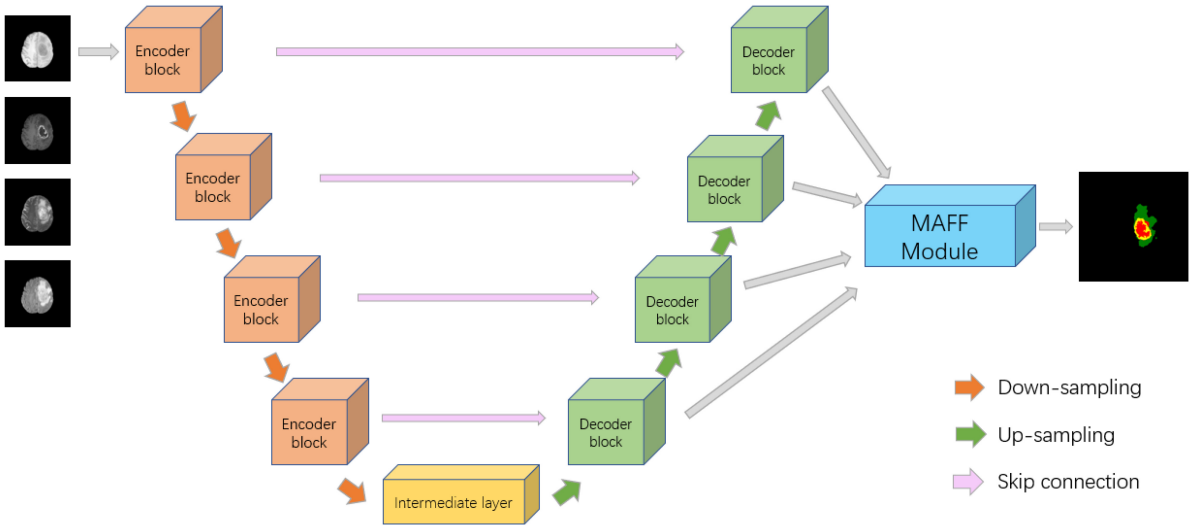


FIGURE 2 | Architecture of the proposed MAFF-ResUNet. In the ground truth image, red, green and yellow stand for tumor core (TC), whole tumor (WT), and enhance tumor (ET), respectively.

different modes have different contrast and other problems, we normalize each modal image before slicing. In this work, z-score normalization is adopted, that is, mean value and standard deviation are used to standardize each image. The formula is as follows:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

where x_i is the input image, and z_i is the normalized image. μ represents the mean value of the input image, while σ denotes the standard deviation of the input image. Finally, we crop the training image into a size of 160 × 160 to reduce the black background in the image, which can obtain effective pixels and reduce the amount of calculation to some extent.

Proposed Method

Architecture of MAFF-ResUNet

Inspired by U-Net (Ronneberger et al., 2015), ResNet (He et al., 2016), DAF (Wang et al., 2018), we propose a deep convolutional neural network with a multi-scale attention feature fusion module based on attention mechanism for brain tumor segmentation. Recently, U-Net has achieved excellent performance in the field of medical image segmentation, which has the advantage of being able to accept input images of any size. **Figure 2** shows the proposed MAFF-ResUNet, which adopts U-Net as our basic network architecture. The MAFF-ResUNet consists of four encoder blocks, four decoder blocks, an intermediate layer and a MAFF module. Firstly, in the down-sampling path, we utilize the convolution layer to extract low-level features of brain tumors, the pooling layer to expand the receptive field, and residual connections to enhance the expression ability of encoder blocks. In the up-sampling path, the up-sampling layer and convolution are used to restore the image resolution. Skip Connections combine low-level information with high-level information to reduce the loss of detailed information.

To further refine the boundaries of the brain tumor, we employ bilinear interpolation to up-sample the feature maps of different resolutions from the four decoder blocks to the same size as the input image and then input them into the MAFF module. The MAFF module extracts attention features of different scales and fuses them to improve the segmentation accuracy of brain tumors and obtain the segmentation results of brain tumors.

Encoder and Decoder Block

Residual U-Net is built by incorporating residual shortcuts into U-Net. Inspired by ResNet (He et al., 2016), we utilize five residual connections in the down-sampling branch, including four encoder blocks and an intermediate layer. The main function of these encoder blocks is to extract low-level features. The introduction of short skip connections is beneficial to obtain better feature expression and accelerate model convergence. As can be seen in **Figure 3A**, each encoder block contains two 3 × 3 convolutions, batch normalization (BN; Ioffe and Szegedy, 2015), Rectified Linear Unit (ReLU) activation function, and a 2 × 2 max-pooling layer. Each Max-pooling layer reduces the size of the input feature map to half of the original. Moreover, the intermediate layer plays the role of connecting the down-sampling and up-sampling paths. Structurally, the intermediate layer is similar to the encoder block, but without the pooling layer.

In the decoding stage, we use four decoder blocks, each of which contains an up-sampling layer, and two 3 × 3 convolutions (see **Figure 3B**). Similarly, each convolutional operation is followed by a BN layer and a ReLU activation layer. The up-sampling layer restores the size of the feature map by using the bilinear interpolation. The input of each decoder block is composed of two parts, one is the output of the previous decoder

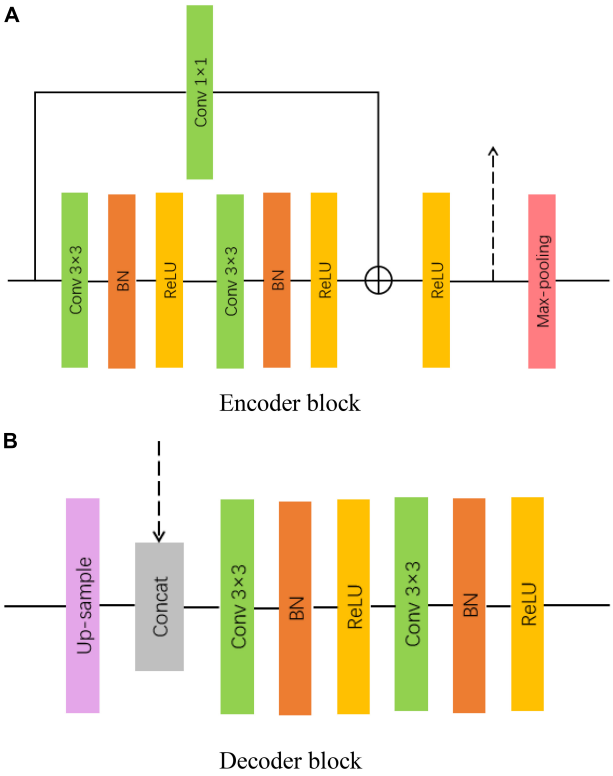


FIGURE 3 | Encoder and decoder block in MAFF-ResUNet **(A)** encoder block **(B)** decoder block.

block, and the other is the output feature map of the same level encoder block, which makes up for the low-level details lost in the high-level semantic space.

Multi-Scale Attention Feature Fusion Module

Inspired by DAF (Wang et al., 2018), we propose a MAFF module to fuse different scale features and improve the accuracy of brain tumor segmentation. As shown in **Figure 4**, the MAFF module accepts feature maps from four different scales, expressed as $F_i \in R^{C \times H \times W}$ $(i = 1, 2, 3, 4)$, where i indicates the feature map of the i-th level, C is the number of channels, H and W represent the height and width of F_i, respectively. These four feature maps are concatenated in the channel dimension and named $F_m \in R^{4C \times H \times W}$. The low-level feature map contains abundant boundary information of brain tumors, while the high-level feature map contains advanced semantic information of brain tumors.

The direct concatenation of different scale feature maps will certainly bring a little noise. The segmentation results obtained by directly exploiting F_m cannot effectively utilize the complementary information of features at different levels. Therefore, we use the attention feature block (AFB) to get the attention map A_i of each level and then multiply it with the mixed feature map F_m to obtain the redefined feature map $F_{ri} \in R^{C \times H \times W}$ of each scale. Specifically, we do the following for F_i:

$$A_i = f_s\left(f_{up}\left(g\left(f_{cat}F_i, f_{1\times1}\left(F_m\right)\right)\right)\right) \quad (2)$$

where $f_{1\times1}$ is a 1×1 convolution followed by BN and ReLU function, f_{cat}, f_{up}, and f_s denote the operations of concatenation, up-sampling, and softmax function, respectively. The $g(x)$ is an attention feature module composed of convolution and average pooling, which can be formulated as:

$$g(x) = f_{1\times1}(f_p(f_{3\times3}(f_p\left(f_{3\times3}\left(x\right)\right)))) \quad (3)$$

where $f_{3\times3}$ is a convolution layer with the filter size of 3×3, and f_p represents the operation of average pooling. Besides, BN and parametric satisfaction linear Unit (PReLU; He et al., 2015) activation function are adopted after each convolution layer in the attention feature block. The PReLU can be obtained by:

$$PReLU\left(x\right) = \max\left(0, x\right) + a * \min(0, x) \quad (4)$$

where a is a learnable parameter.

Then, the output of $g(x)$ is fed to a softmax layer to obtain the attention map after the up-sample. The mathematical expression of the softmax function is given as:

$$Softmax\left(x_j\right) = \frac{e^{x_j}}{\sum_{n=1}^{N} e^{x_n}}, j = 1, 2, \ldots, n \quad (5)$$

According to the A_i, we can selectively extract brain tumor-related feature information from the original feature map by performing a matrix multiplication between A_i and F_m. The output maps are, respectively, concatenated with F_i, and the convolution operation is performed to obtain the redefined feature maps.

In the end, the redefined feature maps of each layer containing both low-level and high-level information are fused, averaged, and then fed to a sigmoid function to obtain the final segmentation result.

Loss Function

In a specific task, the choice of a suitable loss function has a significant influence on the experimental results. The loss function is utilized to express the degree of difference between the predicted value and the label value. During the training process, the model continuously fine-tunes the weight and bias of the network to minimize the loss function value and improve the performance of the model. In this article, for the brain tumor segmentation task, our loss function consists of binary cross-entropy loss (BCE) and Dice loss. The Dice loss mainly applies the Dice coefficient, which is a similarity measurement function. The Dice loss takes the responsibility for the prediction of the brain tumor globally, while the BCE loss is responsible for the classification of each pixel. They can be expressed as:

$$Loss_{BCE}\left(p_k, g_k\right) = -\frac{1}{n}\sum_k\left(g_k * \log\left(p_k\right) + \left(1 - g_k\right) * \log(1 - p_k)\right) \quad (6)$$

$$Loss_{dice}\left(p_k, g_k\right) = 1 - Dice(p_k, g_k) = 1 - \frac{2\left|p_k \cap g_k\right| + \varepsilon}{\left|p_k\right| + \left|g_k\right| + \varepsilon} \quad (7)$$

where n is the number of samples, p_k and g_k denote the prediction of the proposed model and the ground truth,

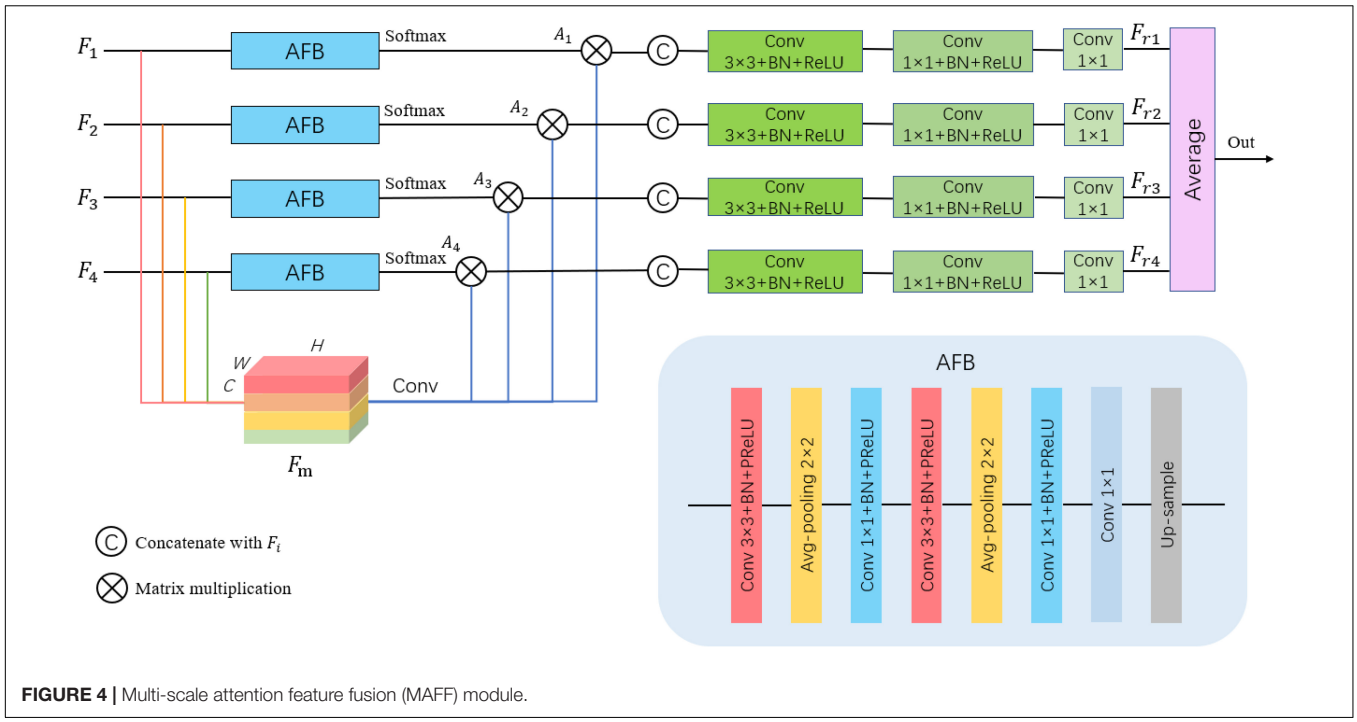


FIGURE 4 | Multi-scale attention feature fusion (MAFF) module.

TABLE 1 | Comparison of segmentation results (mean ± SD) between the proposed MAFF-ResUNet and existing deep convolutional neural networks.

		FCN32s	FCN16s	FCN8s	U-Net	MAFF-ResUNet
IoU (%)	WT	72.9 ± 0.3	77.8 ± 0.4	80.8 ± 0.2	85.6 ± 0.4	**86.5 ± 0.08**
	TC	80.1 ± 0.07	82.1 ± 0.3	84.4 ± 0.07	88.4 ± 0.4	**88.5 ± 0.3**
	ET	66.7 ± 0.6	73.1 ± 0.2	77.7 ± 0.05	85.8 ± 0.4	**86.4 ± 0.3**
SEN (%)	WT	85.0 ± 1.0	87.8 ± 0.8	88.6 ± 0.4	91.1 ± 0.4	**91.93 ± 0.1**
	TC	87.6 ± 0.4	89.7 ± 0.3	90.5 ± 0.05	92.6 ± 0.3	**93.4 ± 0.1**
	ET	75.5 ± 0.5	82.3 ± 0.4	85.0 ± 0.2	91.8 ± 0.3	**92.5 ± 0.1**
PPV (%)	WT	81.7 ± 0.6	85.3 ± 0.3	88.3 ± 0.08	92.0 ± 0.09	**92.5 ± 0.1**
	TC	89.9 ± 0.6	90.1 ± 0.4	92.1 ± 0.2	**94.4 ± 0.06**	93.9 ± 0.3
	ET	82.4 ± 0.3	85.0 ± 0.3	88.7 ± 0.06	92.27 ± 0.7	**92.34 ± 0.4**
DSC (%)	WT	81.9 ± 0.3	85.2 ± 0.3	87.3 ± 0.2	90.5 ± 0.3	**91.2 ± 0.06**
	TC	85.4 ± 0.1	86.8 ± 0.2	88.6 ± 0.1	91.7 ± 0.3	**91.8 ± 0.3**
	ET	73.0 ± 0.6	79.4 ± 0.2	83.4 ± 0.09	89.8 ± 0.4	**90.2 ± 0.3**
HD (mm)	WT	2.92 ± 0.005	2.68 ± 0.01	2.49 ± 0.01	2.20 ± 0.001	**2.16 ± 0.005**
	TC	1.73 ± 0.005	1.64 ± 0.006	1.56 ± 0.003	1.41 ± 0.01	**1.39 ± 0.006**
	ET	1.79 ± 0.01	1.64 ± 0.004	1.49 ± 0.004	1.23 ± 0.02	**1.20 ± 0.005**

Bold indicates the maximum value of IoU, SEN, PPV, DSC, and the minimum value of HD among these methods.

respectively; $|p_k \cap g_k|$ represents the intersection between p_k and g_k; $|p_k|$ and $|g_k|$ are the number of pixels in p_k and g_k, respectively. ε stands for the smoothing coefficient, and the value is set to 1.0×10^{-5}.

The total loss is described as:

$$Loss = \alpha Loss_{BCE} + \beta Loss_{dice} \quad (8)$$

where α and β represent the weight. We empirically set the weight α as 0.5, and the weight β as 1.

EXPERIMENTS AND RESULTS

Training Details

The proposed MAFF-ResUNet is conducted on the PyTorch framework with an NVIDIA GeForce RTX 3090. In this experiment, we use adaptive moment estimation (Adam) (Kingma and Ba, 2014) as the optimizer. The initial learning rate is 0.0003, momentum is 0.90, and weight decay is set to 0.0001. We utilize poly police to decay the learning rate in the progress of training, as employed by Mou et al. (2019) and

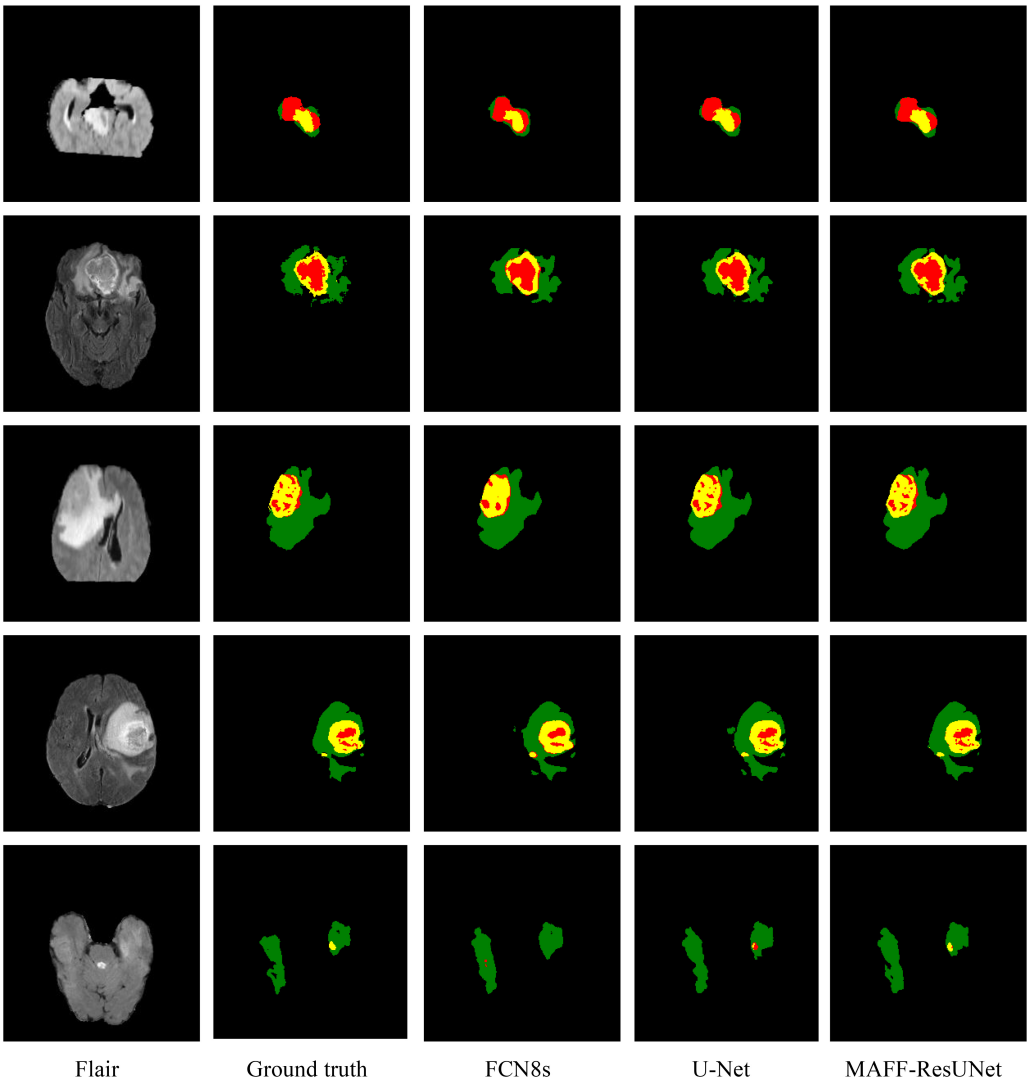


FIGURE 5 | Visualized predicted images of different models. In the ground truth image, red, green and yellow represent tumor core (TC), whole tumor (WT), and enhance tumor (ET), respectively.

Elhassan et al. (2021). It can be defined as in Eq. (9), where *iter* represents the number of iterations, *max_iter* denotes the maximum number of iterations, and *power* is set to 0.9. During training, the batch size is 16.

$$lr = lr_{init} \times (1 - \frac{iter}{max_iter})^{power} \quad (9)$$

Evaluation Metrics

To effectively evaluate the performance of the proposed model, we adopt intersection-over-union (IoU), sensitivity, and positive predictive value (PPV), which are commonly used metrics for image segmentation. The IoU can be calculated using Eq. (10). $(P \cap G)$ is the number of positive pixels which values are the same in both P and G, while $(P \cup G)$ stands for the union of P and G. Sensitivity is defined as the ratio of correctly classified positive samples to the total positive samples in ground truth "G," as shown in Eq. (11). It can be employed to measure the sensitivity of the model to segmentation targets. PPV represents the proportion of correctly classified positive samples to all positive samples in predicted "P," which can be formulated as in Eq. (12). $|P|$ is the number of positive pixels in P, while $|G|$ is the number of positive pixels in G. The IoU, sensitivity, and PPV are all ranging from 0 to 1. The closer they are to 1, the better the segmentation result is:

$$IoU(P, G) = \frac{(P \cap G)}{(P \cup G)} \quad (10)$$

$$SEN(P, G) = \frac{(P \cap G)}{|G|} \quad (11)$$

$$PPV(P, G) = \frac{(P \cap G)}{|P|} \quad (12)$$

Furthermore, as two commonly used metrics for brain tumor segmentation, the Dice similarity coefficient (DSC) and Hausdorff distance (HD) are also applied for the qualitative analysis in this experiment. The DSC is utilized to calculate how similar two samples are, and can be given as:

$$DSC(P, G) = \frac{2 \times |P \cap G|}{|P| + |G|} \quad (13)$$

The DSC is sensitive to the internal padding of the mask. Compared with the DSC, the HD is more sensitive to the segmented boundary. It represents the maximum Hausdorff distance between the labeled boundary and the predicted boundary, defined as:

$$HD(P, G) = \max\left\{d_{PG}, d_{GP}\right\}$$

$$= max\left\{\max_{p\in P}\min_{g\in G}(p, g), \max_{g\in G}\min_{p\in P}(g, p)\right\} \quad (14)$$

where p and g represent the points in the predicted area and the ground truth area, respectively.

Performance Comparison

We compare the proposed MAFF-ResUNet with different networks, including FCN (Long et al., 2015) and U-Net (Ronneberger et al., 2015). For the FCN network, we will use three models with different network depths: FCN8s, FCN16s, FCN32s. As shown in **Table 1**, the various metrics of FCN in the brain tumor segmentation task are lower than that of other models. Compared with FCN, U-Net with encoder-decoder structure has stronger feature extraction capabilities, and the model performance is significantly improved. The MAFF-ResUNet proposed in this article, except for the PPV metric of TC slightly lower than U-Net, other metrics have improved.

Moreover, we evaluate the performance of the proposed model from a more intuitive perspective. **Figure 5** shows the comparison of the prediction image between the proposed approach in this article and other methods. The figure contains four different cases, and shows the original MRI images of flair modality, the prediction results of each model, and ground truth images. Since FCN8s has the best performance among the three FCN networks, we only use the prediction results of FCN8s for comparison. By comparison, we can find that the proposed method in this article is significantly better than U-Net and has obvious advantages in the segmentation of brain tumor contours and edge details. It shows that the introduction of the MAFF module makes the brain tumor segmentation results have richer edge information. Besides, there are fewer pixels mistakenly classified by the MAFF-ResUNet. The predicted images of the MAFF-ResUNet are more similar to manually annotated images.

CONCLUSION

In this article, we propose a deep convolutional neural network for MRI brain tumor segmentation, named MAFF-ResUNet. This network takes advantage of the encoder-decoder structure. The introduction of residual shortcuts in the encoder block, combined with skip connections, enhances the global feature extraction capability of the network. In addition, for the output feature maps of different levels of decoder blocks, the attention mechanism is utilized to selectively extract important feature information of each level. Then the multi-scale feature maps are fused to obtain the segmentation. The proposed method is verified on the public BraTs 2019 MRI dataset. Experimental results show that the MAFF-ResUNet is better than existing deep convolutional neural networks. From the perspective of predicted images, the proposed method can effectively exploit multi-scale feature information and maintain most of the edge detail information. Therefore, the MAFF-ResUNet method proposed in this article can achieve high-precision automatic segmentation of brain tumors and can be used as an auxiliary tool for clinicians to perform early screening or diagnosis and treatment of brain tumors.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.med.upenn.edu/cbica/brats-2019/.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Ethical review and approval was not required for the animal study because the data sets used in this manuscript are from public data sets. The data link is as follows: https://www.med.upenn.edu/cbica/brats-2019/. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XH wrote the main manuscript and conducted the experiments. WX and JY participated in the writing of the manuscript and modified the English grammar of the article. JY and SC made the experiments. JM and ZW analyzed the results. All authors reviewed the manuscript.

FUNDING

This work was supported by a grant from the Natural Science Foundation of Fujian Province (2020J02063), Xiamen Science and Technology Bureau Foundation of Science and Technology Project for Medical and Healthy (3502Z20209005), National Natural Science Foundation of China (82072777), and Xiamen Science and Technology Bureau Foundation of Science and Technology Project for Medical and Healthy (3502Z20214ZD1013).

REFERENCES

Angulakshmi, M., and Deepa, M. (2021). A review on deep learning architecture and methods for MRI brain tumour segmentation. *Curr. Med. Imaging* 17, 695–706. doi: 10.2174/1573405616666210108122048

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., et al. (2017a). Data descriptor: advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data.* 4:13. doi: 10.1038/sdata.2017.117

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., et al. (2017b). Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *Cancer Imaging Arch.* doi: 10.7937/K9/TCIA.2017.KLXWJJ1Q

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., et al. (2018). Identifying the best machine learning algorithms for brain tumor segmentation. progression assessment, and overall survival prediction in the BRATS Challenge. *arXiv* [Preprint] arXiv: 1811.02629,

Bauer, S., Wiest, R., Nolte, L. P., and Reyes, M. (2013). A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* 58, R97–R129. doi: 10.1088/0031-9155/58/13/R97

Dawngliana, M., Deb, D., Handique, M., and Roy, S. (2015). "Automatic brain tumor segmentation in mri: hybridized multilevel thresholding and level set," in *Proceedings of the 2015 International Symposium on Advanced Computing and Communication (ISACC)*, (Silchar: Institute of Electrical and Electronics Engineers), 219–223.

Elhassan, M. A. M., Huang, C. X., Yang, C. H., and Munea, T. L. (2021). DSANet: dilated spatial attention for real-time semantic segmentation in urban street scenes. *Exp. Syst. Appl.* 183:12. doi: 10.1016/j.eswa.2021.115090

Essadike, A., Ouabida, E., and Bouzid, A. (2018). Brain tumor segmentation with vander lugt correlator based active contour. *Comp. Methods Prog. Biomed.* 160, 103–117. doi: 10.1016/j.cmpb.2018.04.004

Gordillo, N., Montseny, E., and Sobrevilla, P. (2013). State of the art survey on MRI brain tumor segmentation. *Magn. Reson. Imaging* 31, 1426–1438. doi: 10.1016/j.mri.2013.05.002

Harati, V., Khayati, R., and Farzan, A. (2011). Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images. *Comput. Biol. Med.* 41, 483–492. doi: 10.1016/j.compbiomed.2011.04.010

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., et al. (2017). Brain tumor segmentation with deep neural networks. *Med. Image Anal.* 35, 18–31. doi: 10.1016/j.media.2016.05.004

He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, (Santiago: Institute of Electrical and Electronics Engineers), 1026–1034.

He, K. M., Zhang, X. Y., Ren, S. Q., 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: Institute of Electrical and Electronics Engineers), 770–778.

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *Proc. Mach. Learn.* 37, 448–456.

Kingma, D., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv* [Preprint] arXiv:1412.6980,

Latif, G., Alghazo, J., Sibai, F. N., Iskandar, D., and Khan, A. H. (2021). Recent advancements in fuzzy c-means based techniques for brain MRI segmentation. *Curr. Med. Imaging* 17, 917–930. doi: 10.2174/1573405616666210104111218

Latif, U., Shahid, A. R., Raza, B., Ziauddin, S., and Khan, M. A. (2021). An end-to-end brain tumor segmentation system using multi-inception-UNET. *Int. J. Imaging Syst. Technol.* 31, 1803–1816. doi: 10.1002/ima.22585

Lee, M., Cho, W., Kim, S., Park, S., and Kim, J. H. (2012). Segmentation of interest region in medical volume images using geometric deformable model. *Comput. Biol. Med.* 42, 523–537. doi: 10.1016/j.compbiomed.2012.01.005

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651.

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., et al. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 114, 97–109. doi: 10.1007/s00401-007-0243-4

Ma, C., Luo, G. N., and Wang, K. Q. (2018). Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images. *IEEE Trans. Med. Imag.* 37, 1943–1954. doi: 10.1109/TMI.2018.2805821

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* 34, 1993–2024. doi: 10.1109/TMI.2014.2377694

Mou, L., Zhao, Y. T., Chen, L., Cheng, J., Gu, Z. W., Hao, H. Y., et al. (2019). CS-Net: channel and spatial attention network for curvilinear structure segmentation. *Med. Image Comput. Comput. Assist. Interv.* 11764, 721–730. doi: 10.1007/978-3-030-32239-7_80

Ostrom, Q. T., Patil, N., Cioffi, G., Waite, K., Kruchko, C., and Barnholtz-Sloan, J. S. (2020). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017. *Neuro Oncol.* 22, iv1-iv96. doi: 10.1093/neuonc/noaa200

Pereira, S., Pinto, A., Alves, V., and Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* 35, 1240–1251. doi: 10.1109/TMI.2016.2538465

Pham, D., Prince, J. L., Xu, C. Y., and Dagher, A. P. (1997). An automated technique for statistical characterization of brain tissues in magnetic resonance imaging. *Intern. J. Pattern Recognit. Artif. Intell.* 11, 1189–1211. doi: 10.1142/S021800149700055X

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. *Intern. Conf. Med. Image Comput. Comput. Assist. Interv.* 9351, 234–241. doi: 10.1007/978-3-319-24574-4_28

Wang, Y., Deng, Z. J., Hu, X. W., Zhu, L., Yang, X., Xu, X. M., et al. (2018). Deep attentional features for prostate segmentation in ultrasound. *Med. Image Comput. Comput. Assist. Interv.* 11073, 523–530. doi: 10.1007/978-3-030-00937-3_60

Xu, C. Y., Pham, D. L., and Prince, J. L. (1997). Finding the brain cortex using fuzzy segmentation, isosurfaces, and deformable surface models. *Intern. Conf. Inf. Process. Med. Imaging* 1230, 399–404. doi: 10.1007/3-540-63046-5_33

Zhang, Z., Wu, C. D., Coleman, S., and Kerr, D. (2020). DENSE-INception U-net for medical image segmentation. *Comput. Methods Prog. Biomed.* 192:105395. doi: 10.1016/j.cmpb.2020.105395

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 30 November 2021
doi: 10.3389/fnins.2021.797990

Electroencephalogram Source Imaging and Brain Network Based Natural Grasps Decoding

Baoguo Xu, Leying Deng, Dalin Zhang, Muhui Xue, Huijun Li, Hong Zeng and Aiguo Song*

The State Key Laboratory of Bioelectronics, Jiangsu Key Laboratory of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing, China

Studying the decoding process of complex grasping movement is of great significance to the field of motor rehabilitation. This study aims to decode five natural reach-and-grasp types using sources of movement-related cortical potential (MRCP) and investigate their difference in cortical signal characteristics and network structures. Electroencephalogram signals were gathered from 40 channels of eight healthy subjects. In an audio cue-based experiment, subjects were instructed to keep no-movement condition or perform five natural reach-and-grasp movements: palmar, pinch, push, twist and plug. We projected MRCP into source space and used average source amplitudes in 24 regions of interest as classification features. Besides, functional connectivity was calculated using phase locking value. Six-class classification results showed that a similar grand average peak performance of 49.35% can be achieved using source features, with only two-thirds of the number of channel features. Besides, source imaging maps and brain networks presented different patterns between each condition. Grasping pattern analysis indicated that the modules in the execution stage focus more on internal communication than in the planning stage. The former stage was related to the parietal lobe, whereas the latter was associated with the frontal lobe. This study demonstrates the superiority and effectiveness of source imaging technology and reveals the spread mechanism and network structure of five natural reach-and-grasp movements. We believe that our work will contribute to the understanding of the generation mechanism of grasping movement and promote a natural and intuitive control of brain–computer interface.

Keywords: natural reach-and-grasp decoding, movement-related cortical potential, EEG source imaging, phase locking value, brain network

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing Medical University, China
Tongguang Ni,
Changzhou University, China

****Correspondence:***
Baoguo Xu
xubaoguo@seu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *19 October 2021*
Accepted: *05 November 2021*
Published: *30 November 2021*

Citation:
Xu B, Deng L, Zhang D, Xue M, Li H, Zeng H and Song A (2021) Electroencephalogram Source Imaging and Brain Network Based Natural Grasps Decoding. Front. Neurosci. 15:797990. doi: 10.3389/fnins.2021.797990

INTRODUCTION

Brain–computer interface (BCI) is a control system, which enables users to directly communicate with the external environment through electroencephalogram (EEG) (Wolpaw et al., 2000). In many applications of BCI, the field of motor rehabilitation has attracted many researchers' attention. Moreover, as an essential human skill, hand function is the first choice for many people with dyskinesia want to restore (Snoek et al., 2004). Studies in the early 21st century have shown

that neural prosthesis control can be achieved by noninvasive BCI, but mainly depends on simple and repeated imagination of motor tasks (Pfurtscheller et al., 2003; Rohm et al., 2013), which gives an unnatural controlling feeling for BCI users.

To extend the BCI instruction set and realize a natural and intuitive control, many researchers have investigated the possibility of decoding complicated grasping information using movement-related cortical potentials (MR) (Jochumsen et al., 2015; Ofner et al., 2016; Pereira et al., 2017; Shiman et al., 2017). MRCP is considered as a low-frequency EEG negative shift that contains rich movement information, including movement types, force, and speed (Shakeel et al., 2015). Using MRCP, Schwarz et al. (2018) successfully decoded three natural grasp types (palmar, pincer, lateral) in the same limb and no-movement condition with a peak performance of 65.9% and identified their significant differences in MRCP. Ofner et al. (2019) investigated MRCPs of five attempted arm and hand movements, and the peak accuracy of five-class classification was 45%. Moreover, they found discriminative signals originated from central motor areas based on pattern analysis. In our previous study, we investigated five natural grasp types and no-movement condition using MRCP, and five-class classification accuracy was significantly better than significance level (Xu et al., 2021).

However, most of these studies are based on channel features. As we know, limited by the volume conductor effect, the spatial resolution of EEG signals is low, which is not conducive for us to understand how EEG is produced and distributed in the brain. To solve this problem, EEG source imaging (ESI) may be a good solution. Many studies about decoding of hand movement EEG reported that the source-based method outperformed sensor-based method (Yuan et al., 2008; Handiru et al., 2016). The previous study by Edelman et al. (2016) verified that ESI can improve decoding performance when classifying four complex right-hand motor imagery (MI) tasks: the accuracy of individual task classification and overall classification after using ESI was, respectively, 18.6 and 12.7% higher than the peak performance obtained using the sensor-based method. Moreover, a recent study that combined scout ESI and convolutional neural network presented a high classification performance for 4-class MI tasks: the overall accuracy was 14.4% higher than the accuracy achieved using the state of the art MI classification methods (Hou et al., 2020). Because of the proximity of motor brain regions activated by the same limb, we hope to improve spatial resolution using ESI technology and obtain a higher classification result.

Recently, a novel and effective way to investigate the communication patterns of the brain regions has attracted many researchers' attention, which is brain network analysis. The study of brain network is often described by functional connectivity and graph theory. For example, the recent study by Gu et al. revealed the network distinction between left and right foot MI tasks based on network-based statistics and graph theory indices (Gu et al., 2020). In addition, Xiong et al. (2020) constructed functional connectivity matrices in five frequency bands based on weighted phase lag index and used the network metrics from these matrices to decode different movement intentions. However, few studies focus on constructing brain networks in MRCP frequency band, especially on natural grasping movements.

In our study, we collected EEG signals for five natural grasp types as well as no-movement condition. We applied ESI technology on the MRCP frequency band and used phase locking value (PLV) between the regions of interest (ROIs) to construct functional connectivity matrices. In addition, five network metrics were calculated to show the changes of brain networks of different movements over time. Our study aimed to answer three questions: (1) whether the classification performance can be improved by using source features compared with using channel features; (2) whether the conclusions obtained by using source features are consistent with those obtained by using channel features; and (3) whether there are significant differences in brain networks and network metrics of different grasp types and what changes they show in the whole reach-and-grasp process. We hope our study can shed some light on other researchers.

MATERIALS AND METHODS

Subjects

This study was approved by the ethics committee of Southeast University. Eight right-handed subjects (five female) aged between 22 and 25 years were recruited. All subjects were without any history of neuromuscular disorders and had normal or corrected-to-normal vision. All subjects were accurately explained the purpose and procedure of the experiment and signed the informed consent.

Experimental Setup

The experiment was carried out in an electromagnetic and noise-shielded room with weak light, good ventilation, and appropriate temperature. We used an audio cue-based paradigm, as shown in **Figure 1A**. Subjects were instructed to perform six kinds of tasks, which were palmar, pinch, push, twist, plug, and no-movement, as illustrated in **Figure 1C**, and these tasks appeared in random order. At second 0, an auditory beep instructed subjects to place

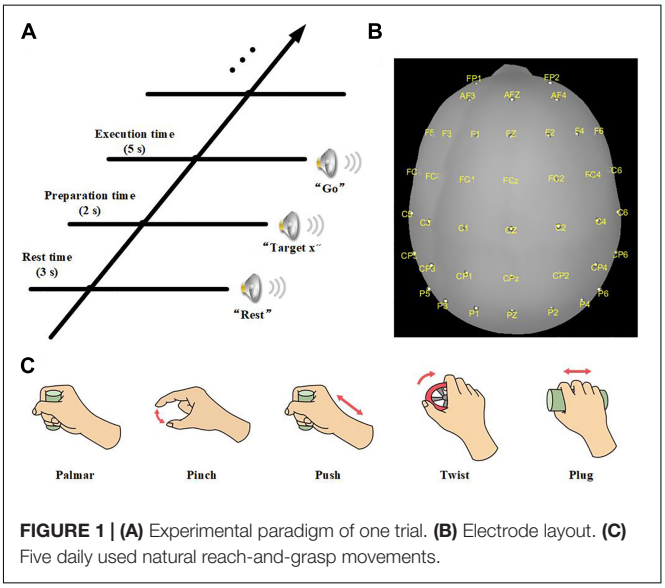


FIGURE 1 | (A) Experimental paradigm of one trial. **(B)** Electrode layout. **(C)** Five daily used natural reach-and-grasp movements.

their right hands on the press button and take a 3-s rest. After that, an auditory cue indicated a specific task, and subjects were asked to stay focused and avoid swallowing or blinking. At second 5, a "go" cue instructed subjects to execute the required movement in 5 s, including reaching, grasping, and returning. For no-movement instructions, subjects should avoid any eye or body movement.

The experiment was divided into 8 sessions, and 10 trials were recorded for each task in a session. Therefore, 480 trials were obtained for each subject. Furthermore, we recorded a 10-s rest at the beginning of the experiment, and an 8-min rest was introduced between sessions to reduce the fatigue of subjects.

Data Recording

We used a 64-channel active electrode cap (BrainProducts, Germany) and a SynAmps2 amplifier (Neuroscan Compumedics, United States) for EEG recording. The electrodes were arranged following the international standard 10–20 montage. Forty electrodes positioned over frontal, parietal, and temporal lobes were selected to collect the signals (**Figure 1B**). We used the left mastoid as a reference and the FPz channel as ground. The electrode impedances were kept less than 5 kΩ. EEG was sampled with 1,000 Hz and prefiltered between 0.05 and 100 Hz. Besides, a notch filter at 50 Hz was applied to remove power-line noises. In addition, we used a press button and force transducers to record force information and analyze the reach-and-grasp phase.

Movement Onset Detection

For detecting the movement onset of each trial, the rising edge of the pressure button was used. Similarly, the falling edge of the pressure button was defined as the end of the movement. For no-movement conditions, we calculated the mean reaction time of movement trials of each subject and added it to the "go" cue onset of no-movement trials. The reaction time was the period between the "go" cue onset and the actual movement onset. In addition, the time when subjects started and finished their grasp was also determined by the rising and falling edge of corresponding force signal. In this step, we discarded the trials that met one of the following four conditions: (1) subject released press button before "go" cue; (2) subject did not return to the press button in 5 s after "go" cue; (3) subject's reaction time was more than 1.5 s; and (4) subject executed a wrong movement.

Signal Preprocessing

Signals were processed using MATLAB R2020b and EEGLAB toolbox (Delorme and Makeig, 2004).

First, EEG signals were filtered from 0.1 to 40 Hz using a zero-phase fourth-order Butterworth filter and then downsampled to 100 Hz. Afterward, noisy channels were removed by visual inspection.

To remove stationary artifacts, we high-passed filtered EEG greater than 1 Hz using a zero-phase fourth-order Butterworth filter and computed independent components (ICs) with the extended infomax independent component analysis (ICA) (Lee et al., 1999). In this step, the ICA weights were cached, and contaminated ICs were visually identified and removed with the help of EEGLAB implementation of ADJUST (Mognon et al., 2011) and ICLabel (Pion-Tonachini et al., 2019). After that, the rest ICs were projected back to the channel space and trials with (1) amplitude exceeding ± 100 μV, (2) abnormal joint probability, and (3) abnormal kurtosis were marked for rejection. The threshold of (2) and (3) was five times the standard deviation of their statistic.

Then, we applied the cached ICA weights to the original data (0.1–40 Hz filtered). Previously identified ICs were rejected, and data were back-projected to the channel space. We extracted epochs in the [−2, 3.5] s interval of each trial as time ROI (tROI), with 0 s corresponding to movement onset. Furthermore, trials with rejection marks were excluded.

Lastly, data were converted to a common average reference. We adopted a zero-phase fourth-order Butterworth filter between 0.1 and 3 Hz to get MRCP.

Electroencephalogram Source Imaging

To make results more explanatory, MRCP were mapped from the channels to the cortical surface through ESI. The ESI processing was based on Matlab Brainstorm (Tadel et al., 2011). Colin27 template boundary element head model was chosen because of the high three-dimensional resolution. We used OpenMEEG (Gramfort et al., 2010) to solve the forward problem and sLORETA (Pascual-Marqui, 2002) to compute sources. To reduce the impact of sensor noise, we used 10-s resting data, which were similarly preprocessed as the movement data, to estimate a noise covariance matrix.

The ROIs in our study were defined by an atlas of a subset of the Brodmann areas (Fischl et al., 2008). There were 24 ROIs, which covered somatosensory area (BA1, BA2 BA3a, BA3b), primary motor area (BA4a, BA4p), premotor area (BA6), Broca area (BA44, BA45), primary visual area (V1), secondary visual area (V2), and visual area of middle temporal (MT) of both hemispheres (**Figure 2**). The purpose of studying the visual areas is to investigate whether different grasping objects have significantly different activation patterns on the visual areas. To reduce the computation, we averaged the sources of MRCP in the same ROI and referred to it as sMRCP.

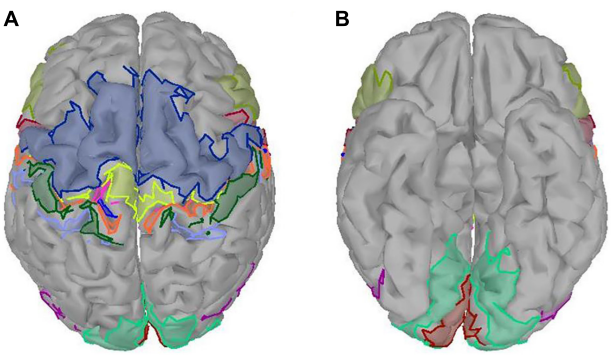


FIGURE 2 | Twenty-four ROIs defined by an atlas of a subset of the Brodmann areas. Different colored regions referred to different ROIs. **(A)** Top view. **(B)** Bottom view.

Feature Extraction and Classification

In our study, we were interested in the six-class classification performance, as in practical application scenarios, BCI equipment should not only distinguish complicated grasp types, but also recognize whether users want to grasp.

The feature extraction method and classifier in our study followed the practice of Schwarz et al. (2018). The difference was that they extracted the amplitude of MRCP as features, while we extracted the amplitude of sMRCP. First, sMRCP was resampled to 16 Hz to ease computational pressure. Next, a 1-s time window was used to take amplitude values in 125-ms steps as features. For classification, we used shrinkage-based linear discriminant analysis classifiers (sLDA) (Blankertz et al., 2011). For every trial, we moved the time window in steps of 1/16 s over the tROI and finally obtained 88 sLDA models over the whole tROI and 192 features for every model. Finally, fivefold cross-validation was repeated 10 times, and the mean accuracy was reported.

To figure out the impact of the volume conductor effect, we compared the classification performance obtained using channel features and source features. We also investigated the impact of time window size on the overall classification performance using four window sizes (one sample, 0.5, 1, and 1.5 s). Furthermore, we calculated precision and confusion matrices at specific time points to see which grasp type had a major impact on the classification performance.

Brain Network Analysis

To investigate the motion coding process, we calculated functional cortical connectivity using PLV (Lachaux et al., 1999). PLV is a measure of synchronization in the time domain and the single-trial formula definition is as follows:

$$\mathrm{PLV}(t) = \frac{1}{N}\left|\sum_{n=1}^{N} \exp\left(j\Delta\varphi(t)\right)\right|$$

where t is the specific time point, N is the number of sample points, and $\Delta\varphi(t)$ is the difference of instantaneous phases between pairs of ROIs at time t. In this article, instantaneous phases were obtained by computing analytic signal using Hilbert transform, and a sliding window of 1 s with 0.1-s step size was adopted for PLV calculation.

A network is usually represented as a graph that consists of nodes and edges. In our study, nodes were 24 ROIs, and the weight of edges was PLV between pairs of ROIs. To avoid spurious connections, we filtered the edges whose PLV was less than 0.65. Besides, we used Gephi to produce functional connectivity maps (Bastian et al., 2009).

In this study, five network metrics were applied to investigate the trends and properties of networks, which were the mean degree (DE), density (DS), characteristic path length (CPL), clustering coefficient (CC), and modularization (M) (Fornito et al., 2016). DE of a node is the sum of the weight of edges connecting this node with all other nodes, which describes the significance of a node. DS is the ratio of the number of actual connections to the maximum number of possible connections. CPL is the average shortest path length between two nodes in a network. CC can be thought of as the probability of finding a connection between any two neighbors of the node. M can be used to divide nodes into densely connected subgroups called modules, and nodes within these modules are more strongly connected than with other parts of the network. CPL indicates functional integration, whereas CC and M indicate functional segregation.

RESULTS

Behavioral Analysis

Figure 3 presents the results of behavioral analysis of all subjects for each reach-and-grasp movement. The movement onset and end were defined by the press button, and the grasping phase was specified by force transducers. The 0 s in the figure corresponds to the subject-specific movement onset. As shown in **Figure 3**, the movement time was different between subjects but keep similar between movement tasks, which pointed out that the more time the subject spent on the palmar task than other subjects, the more time he/she spent on other tasks. For every subject, there was no significant reaction time difference between movement tasks. Besides, we observed similar duration for the reaching phases and returning phases. On average, all subjects completed grasping in 2 s after movement onset. In our study, the pinch-and-twist tasks were completed quicker than the other three tasks, which generally cost 3 s.

Electroencephalogram Source Imaging

Figure 4 shows the grand average of all trials of sMRCPs for each condition in BA2, BA4a, BA6, V1, V2, and BA44, corresponding to six brain regions described above.

In **Figure 4**, we found that in motor-related ROI, for example, BA2, BA4a, and BA6, movement-specific sMRCP in the left hemisphere was more discriminative than that in the right hemisphere, and lateralization could be observed. In addition, we observed a common phenomenon in the left hemisphere of primary motor area (BA4a) and premotor area (BA6): sMRCP often kept steady before 0.7 s and then shifted sharply and reached its peak at approximately 1.5 s, corresponding to the time subjects started grasping. It is worth mentioning that the peak amplitude and peak time of different movement tasks show great distinction. In somatosensory areas, for example, BA2, sMRCP has three peaks, which are at approximately −1.5, 0, and 1 to 1.5 s, respectively. The polarity of these peaks is the same in the left hemisphere, whereas in the right hemisphere, the polarity of the third peak is opposite to that of the first two peaks. Furthermore, the amplitude of the first peak of no-movement is prominently higher than the first peak of movement tasks.

In visual-related ROIs, such as V1 and V2, we did not observe lateralization, but we found significant differences between each grasp type and no-movement condition from 0 to 1 s, corresponding to the reaching phase. Besides, the polarity of sMRCP in the returning phase is opposite to that of the reaching phase. Because of the significant directional sensitivity of the

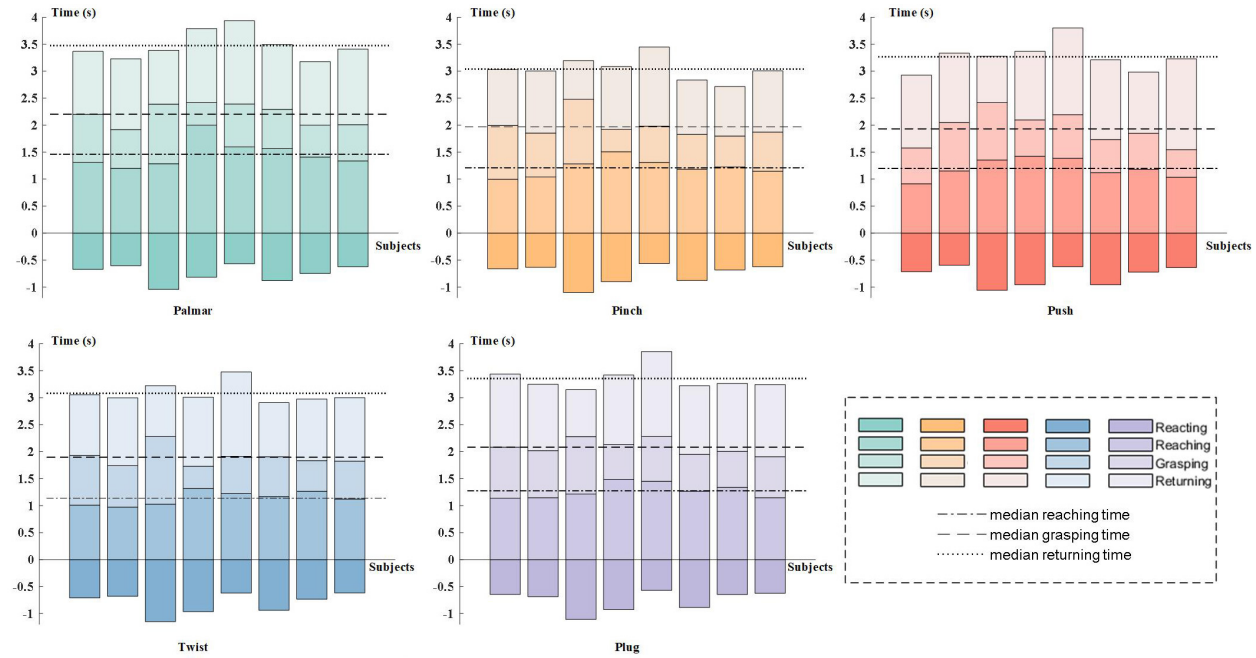


FIGURE 3 | Subject-specific behavioral analysis. In each subgraph, subjects are represented on the horizontal axis in ascending order of subject number from left to right. The 0 s is the subject-specific movement onset. The color tone from dark to light corresponds to the four stages of reach-and-grasp action: reacting phase, reaching phase, grasping phase, and returning phase.

visual areas, we speculated that these phenomena were caused by the object position processing. Moreover, sMRCPs in the left hemisphere of V1 and V2 show a very similar pattern, whereas in the right hemisphere, sMRCP of all conditions in V2 is more discriminative than that in V1 during reaching phase. These phenomena indicated that the same movement activated the left hemisphere of the visual areas in a similar way, whereas in the right hemisphere, secondary visual area (V2) was responsible for more advanced and finer visual processing tasks than the primary visual area (V1).

Figure 5 presents the changes of sMRCP activity of a representative subject for all conditions in the cortex over time. In **Figure 5**, we found that grasping movements mainly activated frontal and parietal areas, and the activation patterns were different. In the time of [−1, 0.5] s, the activation of palmar on the cerebral cortex is the least obvious, mainly focusing on the ipsilateral region. In addition, we observed that two groups had similar activation patterns during [−1, 0.5] s: pinch and push, twist, plug, and no-movement. After 0.5 s, the activation starts to spread from the frontal lobe to the parietal and temporal lobes and then shrinks to the central area. Furthermore, we found that the most pronounced activation was at 1 s in the parietal area, except plug and no-movement, corresponding to the time when the subject reached the object and started grasping. At 2 s, we only found obvious activation in the central area, except push. We speculated this was because this time point corresponded to the start of the returning phase under other conditions, but was still grasping phase for push. In addition, **Figure 5** also indicated that movement intention activated frontal area most, whereas movement execution mainly activated the parietal area.

Multiclass Classification Results

Figure 6A demonstrates the subject-specific and grand average multiclass classification accuracy. The grand average significance level for six-class classification is 17.95% ($\alpha = 0.05$). In **Figure 6A**, the grand average accuracy is above significance level during the whole tROI, which not only verifies that sMRCPs can be used to decode movement intentions of complicated grasp types, but also shows excellent performance of source features. In the beginning, the accuracy curves increase slowly until −1 s. After that, there is a temporary decrease followed by a deep increase, which reaches its peak at the grasping phase. We speculated that the temporary decrease was affected by the "go" cue. Compared with the "Target x" cue, "go" cue was a neutral stimulus, which would increase redundant information and lead to the decrease in accuracy.

Figure 6B presents the grand average precision of each task over tROI. All precision curves have similar trends, which are consistent with **Figure 6A**, whereas the precision of no-movement is significantly higher than that of movement types before the grasping phase. Furthermore, in grasp types, palmar and plug are the easiest to be discriminated, whereas push is the hardest to be identified.

In **Figure 6A**, we found the classification performance varied among subjects. According to their performance, we divided these subjects into two groups. Group 1 consisted of three subjects with better classification performance: S2, S7, and S8, and the rest subjects were included in group 2. **Figures 6C,D**

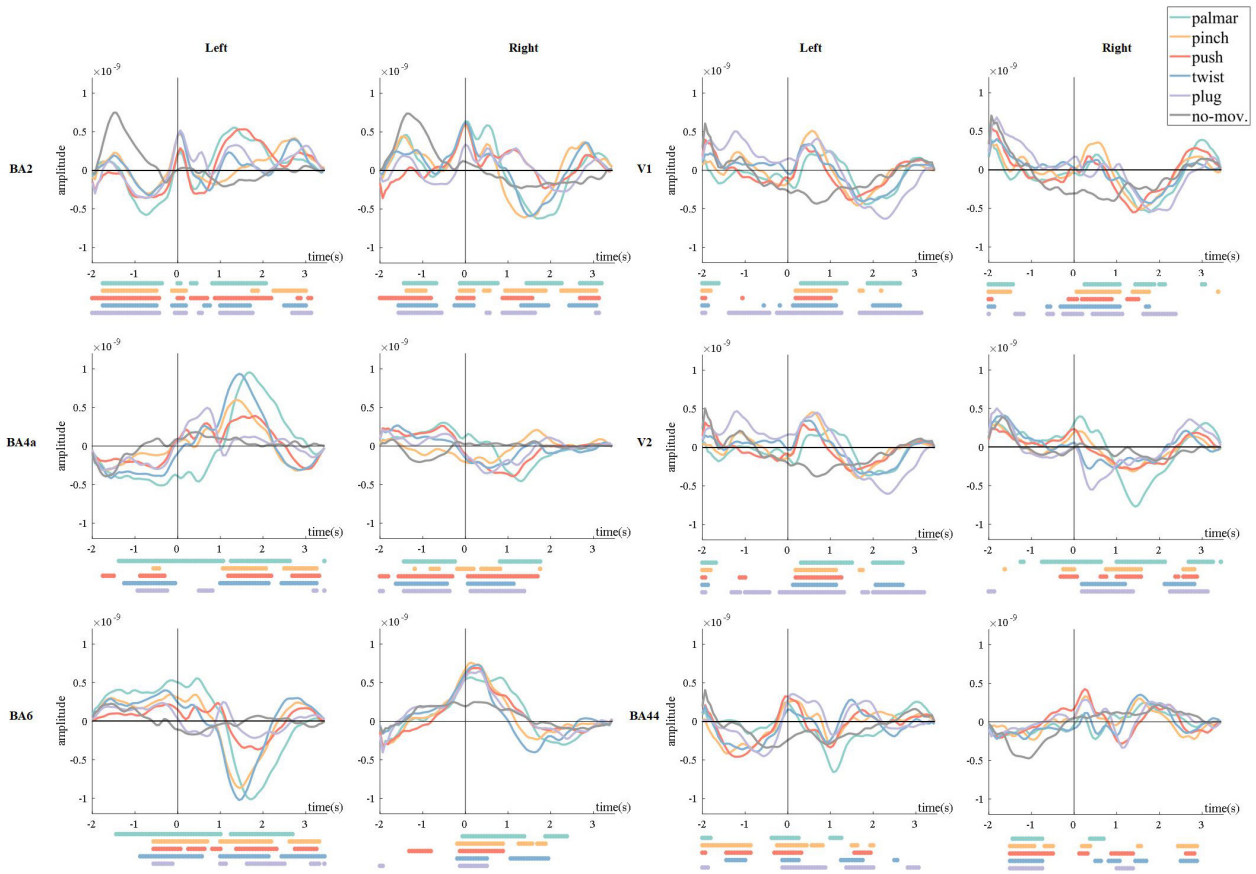


FIGURE 4 | Grand average of all trials of sMRCPs for all conditions. The 0 s corresponds to the movement onset. The dots below each figure were significantly different ($p < 0.05$, Wilcoxon rank sum test) time points between every reach-and-grasp movement and the no-movement.

present confusion matrices of the two groups at grand average peak time. In both groups, no-movement is easily recognized as pinch, whereas pinch is most likely to be falsely identified as twist, followed by push and no-movement. Group 1 has significantly poor performance in push types, whereas the performance of all types seems more similar in group 2.

Table 1 presents the peak performance obtained using source features and channel features. For some subjects, using source features can achieve better performance, whereas for others, using channel features is better. However, grand average performances are almost the same: 49.35% for source features and 49.65% for channel features. Besides, the reduction of features greatly improves the classification efficiency, which is reflected in the execution time: the single trial execution time of the source-based method is only one-third of that of the channel-based method, which indicates the effectiveness and superiority of source-based method.

Besides, the effect of window size on the overall performance was investigated. **Figure 7A** demonstrates the grand average six-class classification performance obtained using four time windows, which is one sample, 0.5, 1, and 1.5 s, respectively. We found that with an increase in the window size, the peak accuracy was higher, and the peak time was later. A one-way repeated-measures analysis of variance was performed on the peak accuracies of all subjects at four window sizes. The results are shown in **Figure 7B**. There is also almost no difference between the performance obtained using a time window of 1 and 1.5 s, but the 1-s window performs significantly better than the 0.5-s window. Therefore, we indicated that the length of source discriminated information was between 1 and 1.5 s.

Brain Network

Figure 8 shows functional connectivity maps of a representative subject (S2) for each condition. **Tables 2–6** present DE, DS, M, CC, and CPL of these networks at −1, 0, and 1.4 s, corresponding to the planning phase, movement onset, and peak time. We observed some interesting phenomena from these networks.

First, the ROI in the same module is almost always in the same hemisphere. For all these networks, we found only one module that contained ROIs in different hemispheres. This module consists of V1, V2, and MT, representing the visual area.

Second, although the number of modules has a little difference between these networks, which often kept 6, sometimes was 5 or 7, reflected in **Table 4**, we found (1) both hemispheres of V1, V2, and MT; (2) left hemisphere of BA44 and BA45;

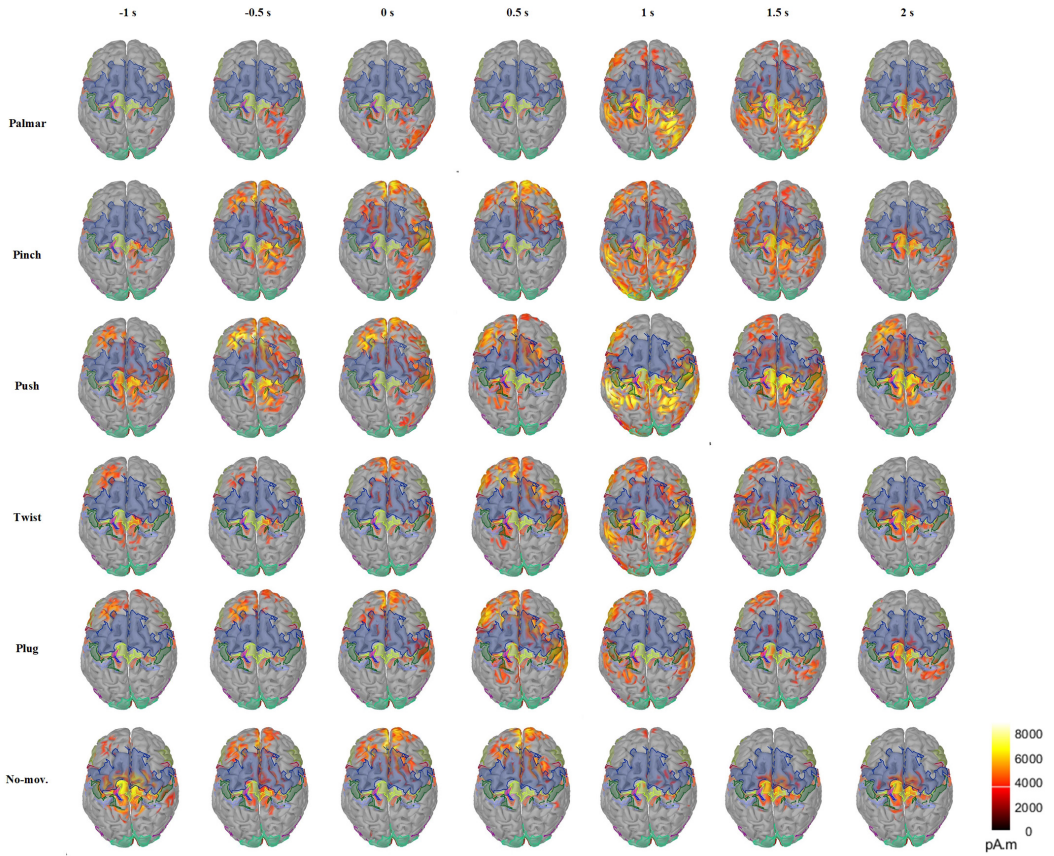


FIGURE 5 | Average cortex maps' changes for all the conditions of S2 in the time of [−1, 2] s, with 0 s corresponding to the movement onset. In this figure, the activity of sMRCP less than 40% of the maximum activation amplitude was filtered.

(3) right hemisphere of BA44 and BA45; (4) right hemisphere of BA2, BA3a, BA3b, and BA4p; (5) left hemisphere of BA2, BA3a, and BA3b; and (6) left hemisphere of BA1, BA4a, and BA4p are always in the same module in **Figure 8**. On the one hand, these divisions are consistent with the functions of these ROIs, which are described in the ESI section. On the other hand, groups 4 to 6 show that the connectivity of the left and right hemispheres is asymmetric, and connectivity in the same hemisphere is stronger than that in different hemispheres.

Third, we observed that with the time increased, connectivity between modules decreased, whereas connectivity in modules increased. **Tables 5**, **6** describe the same phenomenon, where CC increases and CPL decreases over time. This finding indicated that the brain focused more on the communication between modules in the movement intention phase than in the movement execution phase. The work of modules tended to be localized and refined in execution phase.

Lastly, in **Tables 2**, **3**, DE and DS of pinch, push, and twist show a tendency that decreases first and then increases, which is different from palmar, plug, and no-movement: DE and DS of palmar keep increasing over time, whereas le the same properties of plug and no-movement keep decreasing.

DISCUSSION

In this article, we successfully decoded five natural grasp types and no-movement condition using sMRCP. Multiclass classification results showed that for all subjects, accuracies were better than the grand average significance level during the whole tROI, and a grand average peak accuracy of 49.35% could be achieved at grasping phase. Furthermore, by comparing cortex maps, we found movement intention mainly activated the frontal area, whereas movement execution activated the parietal area. Besides, brain network analysis indicated that modules had more communication in the movement intention phase. During the movement execution phase, the work of modules tended to be localized and refined.

Multiclass Classification Results

Figure 6A shows that the peak accuracy is achieved at grasping phase. We believed that this was because the shape of hands showed the most significant difference in grasping phase compared with other phases, and this information was also reflected in sMRCP.

Figure 6B demonstrates that the precision of no-movement type is significantly higher than movement types before 1 s and

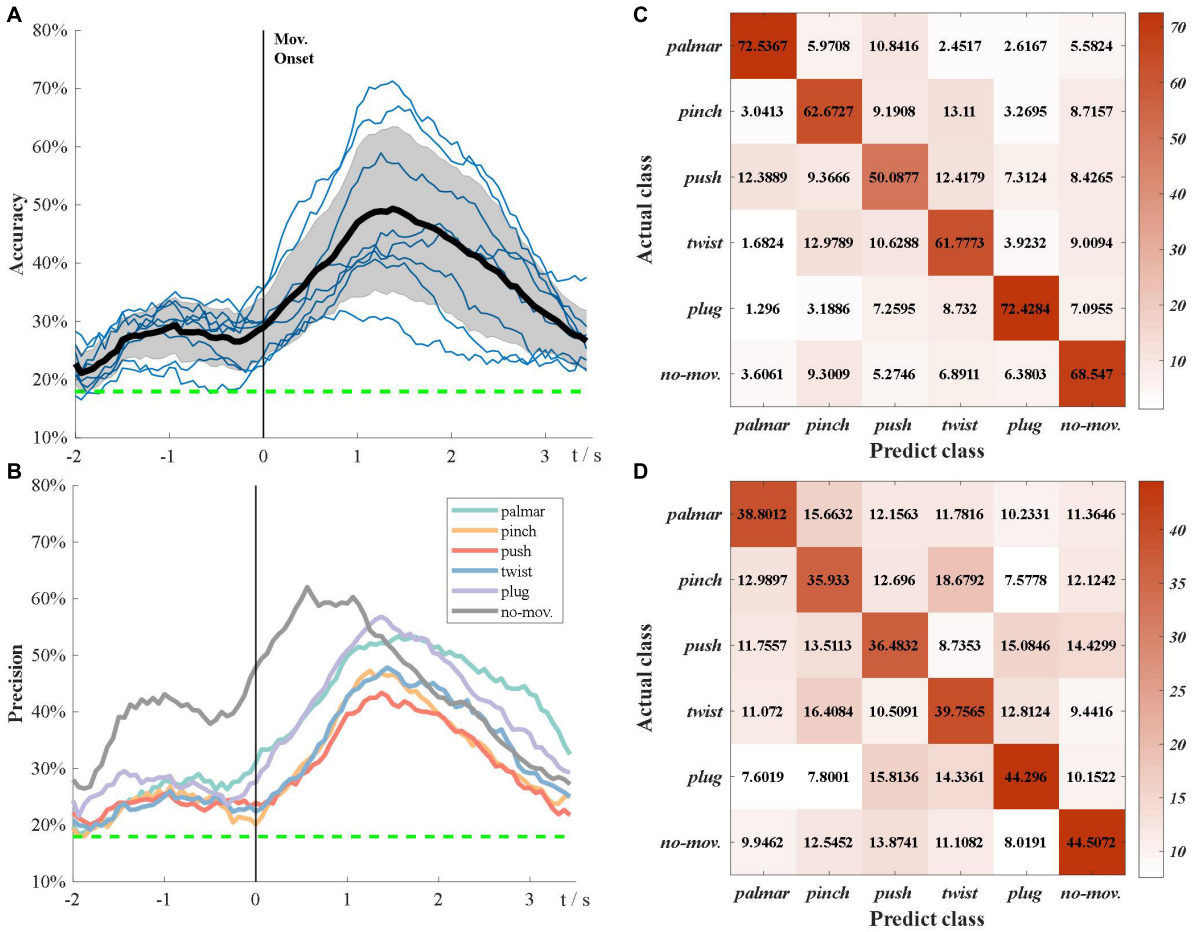


FIGURE 6 | Multiclass classification results. **(A)** Grand average classification accuracy over tROI. The thin blue line represents subject-specific accuracy; the bold line represents grand average accuracy; the shadow shows the standard deviation of grand average accuracy; the green dotted line represents grand average significance level: 17.75%. **(B)** Grand average classification precision of each task over tROI. **(C)** Average confusion matrix of group 1 at grand average peak time. **(D)** Average confusion matrix of group 2 at grand average peak time.

then keeps similar with movement types until the end of the tROI. This phenomenon indicated that significant discriminated information of natural grasp types mainly existed in the grasping phase. In addition, in **Figures 6C,D**, no-movement is easiest to be recognized as pinch, which may be because pinch uses fewer fingers than other grasp types.

Besides, we found that with the window size becoming larger, the peak performance became higher, and the peak time became later. No significant difference was observed between performances obtained using 1- and 1.5-s time window, indicating that the length of discriminative information in sMRCP was between 1 and 1.5 s. These findings are consistent with the study by Schwarz et al., although their study used channel features (Schwarz et al., 2018).

To find out whether source features performed better than channel features, **Table 1** compares the peak performance and the peak time achieved using source features and channel features. Results show that grand average performances are almost the same. It is worth mentioning that the number of source features is less than two-thirds of the number of channel features, indicating that ESI technology is an effective way to improve

TABLE 1 | Peak performance of using source features and channel features and their corresponding time point.

Subject	Source features		Channel features	
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
S1	43.16	1.1875	42.09	1.1250
S2	71.27	1.3750	68.19	1.3125
S3	40.67	1.1875	46.11	1.1250
S4	46.91	1.7500	47.80	1.8125
S5	45.37	1.9375	46.37	1.9375
S6	31.86	0.8125	34.41	0.9375
S7	66.98	1.4375	64.36	1.3750
S8	58.94	1.2500	59.70	1.2500
Grand average	**49.35**	**1.3750**	**49.65**	**1.3750**

The time is relative to movement onset.

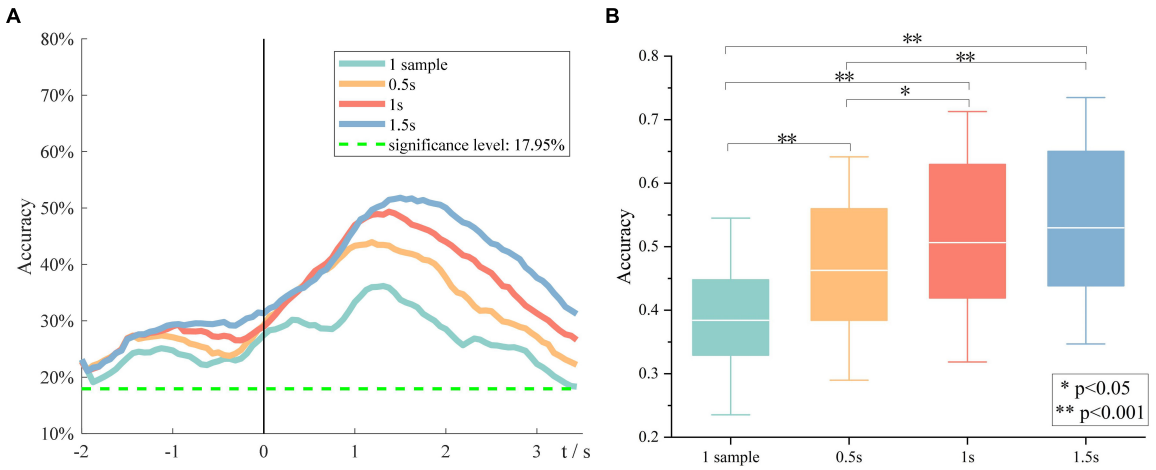


FIGURE 7 | The impact of window size on the overall performance. **(A)** Grand average accuracies of using four window sizes. **(B)** The peak accuracies of all subjects obtained using four window sizes.

TABLE 2 | The average degree of network for each condition of S2.

Time	Palmar	Pinch	Push	Twist	Plug	No-movement
−1 s	4.19	4	5.333	4.182	4.571	5.271
0 s	5.333	3.652	4.167	4.174	4.545	4.545
1.4 s	6.182	5	4.2	4.762	4.2	4

TABLE 3 | The density of network for each condition of S2.

Time	Palmar	Pinch	Push	Twist	Plug	No-movement
−1 s	0.21	0.174	0.267	0.199	0.229	0.237
0 s	0.267	0.166	0.181	0.19	0.261	0.216
1.4 s	0.294	0.263	0.221	0.238	0.221	0.2

TABLE 4 | The modularity of network for each condition of S2.

Time	Palmar	Pinch	Push	Twist	Plug	No-movement
−1 s	0.7	0.753	0.596	0.739	0.601	0.585
0 s	0.61	0.746	0.688	0.772	0.662	0.663
1.4 s	0.594	0.728	0.693	0.7	0.766	0.779

TABLE 5 | The average clustering coefficient of network for each condition of S2.

Time	Palmar	Pinch	Push	Twist	Plug	No-movement
−1 s	0.667	0.625	0.398	0.671	0.585	0.516
0 s	0.71	0.474	0.538	0.729	0.488	0.644
1.4 s	0.745	0.933	0.682	0.786	0.881	0.744

classification performance. Besides, in **Table 1**, compared with the results obtained using channel features, subjects with high performance can achieve better performance after using source features, whereas subjects with a poor performance show poorer results. Moreover, the signal quality also had a great influence on the source imaging results: the better the signal quality, the more accurate the source imaging results.

TABLE 6 | The average path length of network for each condition of S2.

Time	Palmar	Pinch	Push	Twist	Plug	No-movement
−1 s	2.462	2.154	2.309	1.4	2	2.211
0 s	2.026	2.021	2.429	1.513	2.556	2.778
1.4 s	1.976	1.138	2.222	1.545	1.345	1.419

Grasping Pattern Analysis

Previous studies have shown that the discriminative information of grasp types originated from central motor areas (Ofner et al., 2017, 2019). The same conclusion can be obtained from **Figure 5**. Besides, we also observed signals of motor areas were more discriminable in the left hemisphere, and this asymmetry was also presented in the number of the network edges in **Figure 8**. These phenomena reflected the contralateral control of the body by the brain. In addition, we found that the highest distinction of grasp types was in grasping phase in motor and sensorimotor areas. Combined with the research of Sburlea et al. (2021), we speculated that signals in this phase contained not only grasp-type information but also the number of fingers.

Besides, cortex maps in **Figure 5** show a pattern that grasping movements mainly activate the frontal-parietal brain networks, which verifies the results in previous studies. Furthermore, in **Figure 5**, we found that the frontal area was associated with movement intention, whereas the parietal area correlated with movement execution. This is in line with the finding of Sburlea et al. (2021): frontal-central and parietal-occipital activations represent object properties at the planning and execution stage, whereas the primary somatosensory, motor, and parietal areas focus on the processing of grasp types and number of fingers.

In addition, visual area shows a different pattern from motor areas. On the one hand, sMRCPs of motor-related ROIs in **Figure 4** presented significant lateralization, but we did not observe the same phenomenon in visual-related ROIs.

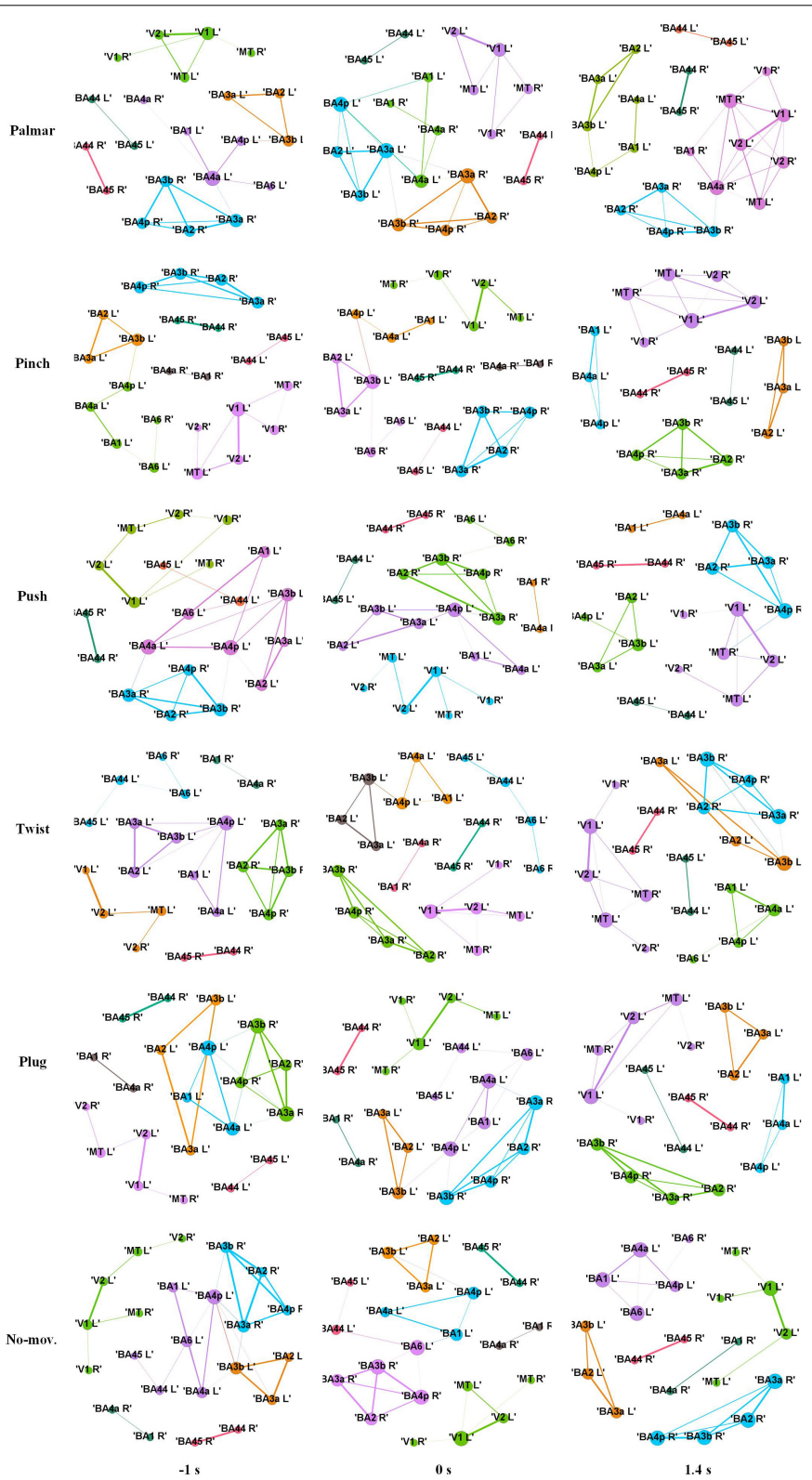


FIGURE 8 | Functional connectivity maps of S2 for each condition. The size of a node indicates DE of the node. The width of an edge indicates the weight of the edge. The edges whose weight is less than 0.65 are filtered. Different colors indicate different modules. Each column represents network structure at different time points, which from left to right are −1, 0, and 1.4 s, respectively.

However, significant differences between each grasp type and no-movement condition were found in the reaching phase. We speculated that visual areas were mainly responsible for processing the object shape and position at this stage. On the other hand, in **Figure 8**, only visual modules contain ROIs of both hemispheres. For motor modules, ROIs of different hemispheres are divided into different modules. This suggested that connections in motor regions are more regional and localized than in visual regions.

Limitations and Future Work

Our study proved that multiclass natural grasp types can be discriminated by MRCP in source space, and source features could achieve similar peak performance with channel features. Besides, grasping pattern was also analyzed using cortex maps and network measures. However, our work still has some limitations.

On the one hand, we used the sLORETA to project MRCP into the source level. Although sLORETA is a mature and popular technology for solving the inverse problem, it is hard to obtain the real source as it is an ill-posed problem. Nonetheless, because of the perfect first-order localization of sLORETA, the source obtained in this study is considered effective.

On the other hand, during brain network analysis, we set a threshold of 0.65 for the PLV connectivity matrices for avoiding spurious connections. We were not sure whether using different thresholds could obtain similar conclusions. For future work, we would compare the impact of using different thresholds.

Besides, the EEG data were collected in healthy people. Since spinal cord injured (SCI) people have dyskinesia, the conclusions obtained from the study of healthy subjects are not necessarily applicable to them. In the future, we will verify our findings on SCI subjects.

CONCLUSION

In our study, five natural reach-and-grasp movements and no-movement were decoded from source amplitude of MRCPs. Compared with using channel features, similar and satisfactory grand average peak performance can be achieved using ESI technology, with only two-thirds of the channel feature number. Grasping pattern analysis demonstrated that the generation patterns varied among different grasp types. Besides, we found that frontal area was associated with movement intention, whereas parietal area correlated with movement execution. Multiclass classification results proved that hand movements are the most distinguishable at grasping stage. Moreover, during the transition from movement intention to movement execution, the work of modules tended to be localized and refined. These findings will contribute to the understanding of the grasping process and eventually realize a natural and intuitive control of BCI.

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 Ethics Committee of Southeast University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BX and LD designed the study, analyzed the data, and wrote the manuscript. DZ set up the experiment platform. MX performed the experiment. HL, HZ, and AS reviewed and edited the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was partially supported by grants from the Basic Research Project of Leading Technology of Jiangsu Province (No. BK20192004) and the National Natural Science Foundation of China (Nos. 62173088 and 62173089).

ACKNOWLEDGMENTS

We would like to thank Yong Wang, Xin Wang and numerous helpers for their suggestions in implementing experiments.

REFERENCES

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an Open Source Software for Exploring and Manipulating Networks. *Proc. Int. AAAI Conf. Web Soc. Media* 3, 361–362.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components–a tutorial. *NeuroImage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

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, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Edelman, B. J., Baxter, B., and He, B. (2016). EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. *IEEE Trans. Biomed. Eng.* 63, 4–14. doi: 10.1109/TBME.2015.2467312

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T. T., et al. (2008). Cortical folding patterns and predicting cytoarchitecture. *Cereb. Cortex* 18, 1973–1980. doi: 10.1093/cercor/bhm225

Fornito, A., Zalesky, A., and Bullmore, E. T. (2016). *Fundamentals Of Brain Network Analysis*. Amsterdam: Academic Press.

Gramfort, A., Papadopoulo, T., Olivi, E., and Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomed. Eng. Online* 9:45. doi: 10.1186/1475-925X-9-45

Gu, L., Yu, Z., Ma, T., Wang, H., Li, Z., and Fan, H. (2020). EEG-based Classification of Lower Limb Motor Imagery with Brain Network Analysis. *Neuroscience* 436, 93–109. doi: 10.1016/j.neuroscience.2020.04.006

Handiru, V. S., Vinod, A. P., and Guan, C. (2016). Multi-direction hand movement classification using EEG-based source space analysis. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2016, 4551–4554. doi: 10.1109/EMBC.2016.7591740

Hou, Y., Zhou, L., Jia, S., and Lun, X. (2020). A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN. *J. Neural Eng.* 17:16048. doi: 10.1088/1741-2552/ab4af6

Jochumsen, M., Niazi, I. K., Taylor, D., Farina, D., and Dremstrup, K. (2015). Detecting and classifying movement-related cortical potentials associated with hand movements in healthy subjects and stroke patients from single-electrode, single-trial EEG. *J. Neural Eng.* 12:56013. doi: 10.1088/1741-2560/12/5/056013

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208. doi: 10.1002/(sici)1097-019319998:4<194:aid-hbm4<3.0.co;2-c

Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719

Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2011). ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240. doi: 10.1111/j.1469-8986.2010.01061.x

Ofner, P., Schwarz, A., Pereira, J., and Müller-Putz, G. (2016). "Movements of the same upper limb can be classified from low-frequency time-domain EEG signals," in *Proceedings of the Sixth International Brain-Computer Interface Meeting: BCI Past, Present, and Future*, (Austria: Verlag der Technischen Universität Graz), 69.

Ofner, P., Schwarz, A., Pereira, J., and Müller-Putz, G. R. (2017). Upper limb movements can be decoded from the time-domain of low-frequency EEG. *PLoS One* 12:e0182578. doi: 10.1371/journal.pone.0182578

Ofner, P., Schwarz, A., Pereira, J., Wyss, D., Wildburger, R., and Müller-Putz, G. R. (2019). Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury. *Sci. Rep.* 9:7134. doi: 10.1038/s41598-019-43594-9

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24, 5–12.

Pereira, J., Ofner, P., Schwarz, A., Sburlea, A. I., and Müller-Putz, G. R. (2017). EEG neural correlates of goal-directed movement intention. *NeuroImage* 149, 129–140. doi: 10.1016/j.neuroimage.2017.01.030

Pfurtscheller, G., Müller, G. R., Pfurtscheller, J., Gerner, H. J., and Rupp, R. (2003). 'Thought' – control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neurosci. Lett.* 351, 33–36. doi: 10.1016/S0304-3940(03)00947-9

Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* 198, 181–197. doi: 10.1016/j.neuroimage.2019.05.026

Rohm, M., Schneiders, M., Müller, C., Kreilinger, A., Kaiser, V., Müller-Putz, G. R., et al. (2013). Hybrid brain-computer interfaces and hybrid neuroprostheses for restoration of upper limb functions in individuals with high-level spinal cord injury. *Artif. Intell. Med.* 59, 133–142. doi: 10.1016/j.artmed.2013.07.004

Sburlea, A. I., Wilding, M., and Müller-Putz, G. R. (2021). Disentangling human grasping type from the object's intrinsic properties using low-frequency EEG signals. *Neuroimage Rep.* 1:100012. doi: 10.1016/j.ynirp.2021.100012

Schwarz, A., Ofner, P., Pereira, J., Sburlea, A. I., and Müller-Putz, G. R. (2018). Decoding natural reach-and-grasp actions from human EEG. *J. Neural Eng.* 15:16005. doi: 10.1088/1741-2552/aa8911

Shakeel, A., Navid, M. S., Anwar, M. N., Mazhar, S., Jochumsen, M., and Niazi, I. K. (2015). A Review of Techniques for Detection of Movement Intention Using Movement-Related Cortical Potentials. *Comput. Math. Methods Med.* 2015:346217. doi: 10.1155/2015/346217

Shiman, F., López-Larraz, E., Sarasola-Sanz, A., Irastorza-Landa, N., Spüler, M., Birbaumer, N., et al. (2017). Classification of different reaching movements from the same limb using EEG. *J. Neural Eng.* 14:46018. doi: 10.1088/1741-2552/aa70d2

Snoek, G. J., IJzerman, M. J., Hermens, H. J., Maxwell, D., and Biering-Sorensen, F. (2004). Survey of the needs of patients with spinal cord injury: impact and priority for improvement in hand function in tetraplegics. *Spinal Cord* 42, 526–532. doi: 10.1038/sj.sc.3101638

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/TRE.2000.847807

Xiong, X., Yu, Z., Ma, T., Luo, N., Wang, H., Lu, X., et al. (2020). Weighted Brain Network Metrics for Decoding Action Intention Understanding Based on EEG. *Front. Hum. Neurosci.* 14:232. doi: 10.3389/fnhum.2020.00232

Xu, B., Zhang, D., Wang, Y., Deng, L., Wang, X., Wu, C., et al. (2021). Decoding Different Reach-and-Grasp Movements Using Noninvasive Electroencephalogram. *Front. Neurosci.* 15:1206. doi: 10.3389/fnins.2021.684547

Yuan, H., Doud, A., Gururajan, A., and He, B. (2008). Cortical imaging of event-related (de)synchronization during online control of brain-computer interface using minimum-norm estimates in frequency domain. *IEEE Trans. Neural Syst. Rehabil. Eng.* 16, 425–431. doi: 10.1109/TNSRE.2008.2003384

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 December 2021
doi: 10.3389/fnins.2021.765634

Quantitative Susceptibility Mapping-Derived Radiomic Features in Discriminating Multiple Sclerosis From Neuromyelitis Optica Spectrum Disorder

*Zichun Yan[1], Huan Liu[2], Xiaoya Chen[1], Qiao Zheng[1], Chun Zeng[1], Yineng Zheng[1], Shuang Ding[1], Yuling Peng[1] and Yongmei Li[1]**

[1] Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2] GE Healthcare, Shanghai, China

Objectives: To implement a machine learning model using radiomic features extracted from quantitative susceptibility mapping (QSM) in discriminating multiple sclerosis (MS) from neuromyelitis optica spectrum disorder (NMOSD).

Materials and Methods: Forty-seven patients with MS (mean age = 40.00 ± 13.72 years) and 36 patients with NMOSD (mean age = 42.14 ± 12.34 years) who underwent enhanced gradient-echo T_2*-weighted angiography (ESWAN) sequence in 3.0-T MRI were included between April 2017 and October 2019. QSM images were reconstructed from ESWAN, and QSM-derived radiomic features were obtained from seven regions of interest (ROIs), including bilateral putamen, globus pallidus, head of the caudate nucleus, thalamus, substantia nigra, red nucleus, and dentate nucleus. A machine learning model (logistic regression) was applied to classify MS and NMOSD, which combined radiomic signatures and demographic information to assess the classification accuracy using the area under the receiver operating characteristic (ROC) curve (AUC).

Results: The radiomics-only models showed better discrimination performance in almost all deep gray matter (DGM) regions than the demographic information-only model, with the highest AUC in DN of 0.902 (95% CI: 0.840–0.955). Moreover, the hybrid model combining radiomic signatures and demographic information showed the highest discrimination performance which achieved the AUC of 0.927 (95% CI: 0.871–0.984) with fivefold cross-validation.

Conclusion: The hybrid model based on QSM and powered with machine learning has the potential to discriminate MS from NMOSD.

Keywords: multiple sclerosis, neuromyelitis optica spectrum disorder, quantitative susceptibility mapping, radiomics, discrimination

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Paola Feraco, University of Bologna, Italy
Gunther Helms, Lund University, Sweden

****Correspondence:***
Yongmei Li
lymzhang70@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *27 August 2021*
Accepted: *01 November 2021*
Published: *03 December 2021*

Citation:
Yan Z, Liu H, Chen X, Zheng Q, Zeng C, Zheng Y, Ding S, Peng Y and Li Y (2021) Quantitative Susceptibility Mapping-Derived Radiomic Features in Discriminating Multiple Sclerosis From Neuromyelitis Optica Spectrum Disorder. Front. Neurosci. 15:765634. doi: 10.3389/fnins.2021.765634

Abbreviations: MS, Multiple sclerosis; NMOSD, Neuromyelitis optica spectrum disorder; QSM, Quantitative Susceptibility Mapping; PUT, Putamen; GP, Globus pallidus; HCN, Head of the caudate nucleus; THA, Thalamus; SN, Substantia nigra; RN, Red nucleus; DN, Dentate nucleus; ROC, Receiver operating characteristic; AUC, Area under the receiver operating characteristic curve; MRI, Magnetic resonance imaging; EDSS, Expanded disability status scale; DGM, Deep gray matter.

INTRODUCTION

Multiple sclerosis (MS) and neuromyelitis optica spectrum disorder (NMOSD) are two major inflammatory demyelinating diseases of the central nervous system (Noseworthy et al., 2000; Wingerchuk et al., 2007). The discovery of the aquaporin-4 antibody (AQP4-Ab) indicates that NMOSD is an inflammatory demyelinating disease independent of MS (Sinnecker et al., 2012; Jurynczyk et al., 2015; Liu et al., 2015; Wingerchuk et al., 2015). NMOSD shows a high prevalence for Asian people. MS and NMOSD share similar clinical manifestations and image features because they both have damages in the optic nerve, spinal cord, and brain, making it difficult to distinguish each other in clinic practice (Huh et al., 2014). Although AQP4-Ab is a common contributor for NMOSD, some NMOSD patients may show a negative autoantibody in cerebrospinal fluid and serum specimens, resulting in incorrect clinical diagnosis and mistreatment (Lennon et al., 2004; Calabrese et al., 2012; Jarius and Wildemann, 2013). In particular, NMOSD sometimes is more serious than MS, and the recommended treatment for MS could exacerbate NMOSD progression (Palace et al., 2010; Jarius and Wildemann, 2013). Therefore, it is vital to diagnose between these two diseases accurately.

Magnetic resonance imaging (MRI) is considered a highly sensitive imaging modality in identifying MS and NMOSD patients. However, conventional MRI scans could not meet the need for higher efficiency of differential diagnosis due to atypical lesion distribution in MS and NMOSD. Therefore, advanced neuroimaging techniques have been used to find more disease characterization and improve optimal therapies. Iron is primarily found in the deep gray matter (DGM) structures within the brain, including globus pallidus (GP), putamen (PUT), head of caudate nucleus (HCN), red nucleus (RN), substantia nigra (SN), thalamic (THA), and the dentate nucleus (DN) (Haacke et al., 2005; Chen et al., 2012). Particularly, previous studies have demonstrated that quantitative susceptibility mapping (QSM), as a novel image sequence, can help to distinguish MS or NMOSD patients from healthy individuals due to the abnormal patterns of iron deposition (Langkammer et al., 2013; Cobzas et al., 2015; Doring et al., 2016; Hagemeier et al., 2017; Zivadinov et al., 2018), which indicated that iron deposition was important in MS and NMOSD pathology. The different QSM-derived magnetic susceptibility values in the DGM between MS and NMOSD were reported by Pudlac et al. (2020). Therefore, QSM may serve as a useful imaging biomarker in guiding early identification and differentiation between MS and NMOSD. However, the mere magnetic susceptibility of the tissue does not ideally account for its iron content because iron occurs in forms of different magnetic properties, and related studies are still limited.

Recently, radiomics extracts the abundant high-throughput radiomic features from conventional medical images and has been recognized as an emerging and attractive research technology, which has shown remarkable success in characterizing tumor phenotypes (Yip and Aerts, 2016). Using hand-crafted features in the regions of interest (ROIs), radiomics makes it possible to extract high-dimensional and mineable features, which can be used for pathophysiology classification in clinical practice (Lambin et al., 2012; Gillies et al., 2016; Yip and Aerts, 2016). Therefore, the QSM sequence was applied to extract radiomic features from DGM regions. Then, a discrimination machine learning model that incorporates the QSM-derived radiomic features and demographic information was performed to classify between MS and NMOSD.

MATERIALS AND METHODS

Standard Protocol Approvals, Registrations, and Patient Consents

This retrospective study has been approved by the Institutional Review Board of the First Affiliated Hospital of Chongqing Medical University, Chongqing, China, and written informed consent was obtained from each participant before MRI scans.

Participants

Patients diagnosed with MS or NMOSD were enrolled from the Department of Radiology, the First Affiliated Hospital of Chongqing Medical University, between April 2017 and October 2019 (**Figure 1**). Patients who met the following inclusion criteria were chosen: (1) patients have received a confirmed diagnosis of NMOSD based on the standard diagnosis criteria (Wingerchuk et al., 2015) and confirmed MS according to the 2017 McDonald criteria (Thompson et al., 2018); (2) patients have undergone enhanced gradient-echo T_2*-weighted angiography (ESWAN) sequence imaging on the same 3.0-T scanner with standardized study protocol; and (3) patients have been in remission (relapse-free for at least 4 weeks) and have no treatment of disease-modified medications within 4 weeks before MRI scans. The exclusion criteria were (1) significant neurologic disease other than MS or NMOSD; (2) image artifacts or incomplete clinical information; and (3) contraindications for MRI scans.

Demographic and clinical data including patient age, sex, disease duration, and Expanded Disability Status Scale (EDSS) score were recorded.

Magnetic Resonance Imaging Acquisitions

All patients underwent an MRI scan of the brain on a 3.0-T system (GE Medical Systems, Milwaukee, WI, United States) using an eight-channel phased-array head coil. The standard protocol for MS studies was performed including a conventional axial 2D dual-echo proton density (PD)-T_2-weighted imaging (T_2WI) [repetition time (TR) = 2,900 ms, echo time (TE)1 = 25 ms, TE2 = 93 ms, echo train length [ETL) = 12, matrix size = 256 × 192] and 2D fluid-attenuated inversion recovery (FLAIR) (TR = 2,050 ms, TE = 24 ms, TI = 750 ms, matrix size = 256 × 256). All axial scans were taken from the roof of the skull to the foramen magnum [slice thickness = 5 mm, slice skip = 0 mm, field of view (FOV) = 24 cm × 24 cm].

ESWAN data were acquired with eight echoes using the following parameters: TR = 60 ms, TE = 6 ms, number of

excitation (NEX) = 0.75, FOV = 22 cm × 22 cm, matrix size = 448 × 320, receiver bandwidth = + 62.5 kHz, and flip angle = 20°. The sequence was acquired with 2-mm-thick contiguous sections and no space. All scans were oriented parallel to the anterior–posterior commissural (AC-PC) line with 56–64 locations on the middle sagittal plane and covered the entire brain area.

Quantitative Susceptibility Mapping Reconstructions

QSM reconstructions were performed using a MATLAB R2013a-based susceptibility imaging software (STISuite[1]). The

[1]https://people.eecs.berkeley.edu/~chunlei.liu/software.html

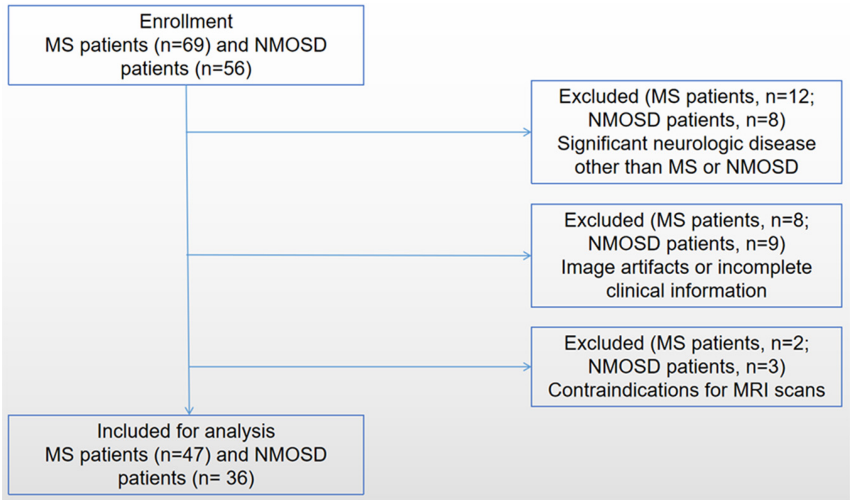


FIGURE 1 | The flowchart shows the enrollment of MS and NMOSD patients.

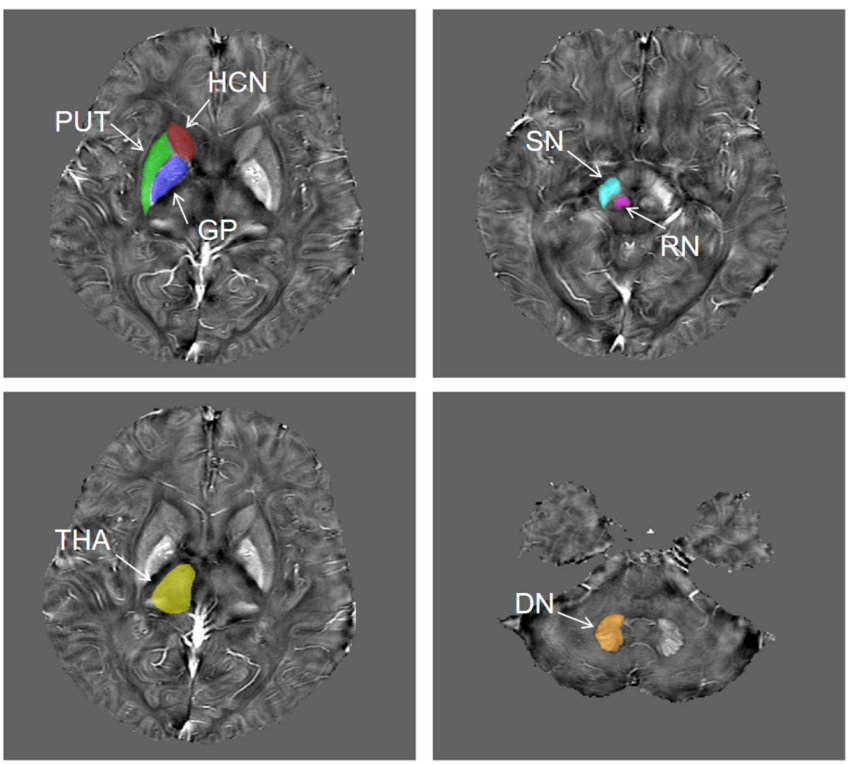


FIGURE 2 | The representative unilateral outlines of the defined ROIs. The ROIs included the putamen (PUT), globus pallidus (GP), head of the caudate nucleus (HCN), thalamus (THA), substantia nigra (SN), red nucleus (RN), and dentate nucleus (DN) bilaterally.

corrected and combined phase images were acquired by weighting the magnitude of the corresponding channel with the vendor-provided combination method and were unwrapped using a Laplacian-phase method. Then, phase-unwrapped images were used to remove the background field using the V-SHARP method (Schweser et al., 2011). In order to reduce extreme streaking artifacts caused by large veins, susceptibility maps were generated in the process of field-to-susceptibility inversion by using an improved sparse linear equation and least-square algorithm (streaking artifact reduction for QSM, STAR-QSM).

Quantitative Susceptibility Mapping Atlas Registration and Regions of Interest Segmentation

The QSM atlas is available in the STISuite V3.0 software package and can be downloaded at, (see text footnote 1) (Zhang et al., 2018). The original QSM data for each participant were registered to the QSM atlas by using the FMRIB Software Library (FSL[2]). Furthermore, seven ROIs for each patient, including bilateral PUT, GP, HCN, THA, SN, RN, and DN, were segmented on QSM images by using Statistical Parametric Mapping (SPM) software based on MATLAB R2013a. **Figure 2** shows the representative unilateral outlines of the defined ROIs.

Feature Extraction and Selection

Before the feature extraction, an isotropic voxel was resampled into 1 mm × 1 mm × 1 mm with linear interpolation for the purpose of normalizing the geometry of MR images. A total of 874 radiomic features were extracted by using the in-house software AK (Artificial Intelligence Kit, version 3.3.0, GE Healthcare, Chicago, IL, United States), including four different sets of features: (1) shape features, (2) first-order features, (3) textural features, and (4) transform features including wavelet and Laplace of Gaussian. The shape-based features were measured using the shape descriptors of three-dimensional size and shape of the ROI. First-order statistical features were applied to describe the distribution of voxel intensities. Texture features used a matrix to represent the spatial heterogeneity of the intensity level. To further investigate the intra-ROI heterogeneity, wavelet filters and Laplace transformation were applied to the original images to convert original images to versions focused on information at different scales. Most of the extracted features obey the imaging biomarker standardization initiative (Zwanenburg et al., 2020). Before the feature selection, the abnormal or missing data were replaced by the median, and the standardization was performed to eliminate dimension differences. Firstly, the univariate analysis was applied to choose the significant features with a *p*-value less than 0.05. Next, the *spearman* correlation analysis between features was used with the cutoff of 0.9, reducing the features redundancy. Then, the multivariate logistic analysis with the Akaike information criterion was used to select the most representative features. Finally, the multivariable logistic regression model with fivefold cross-validation was established. The flowchart of the processing steps is shown in **Figure 3**.

[2]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

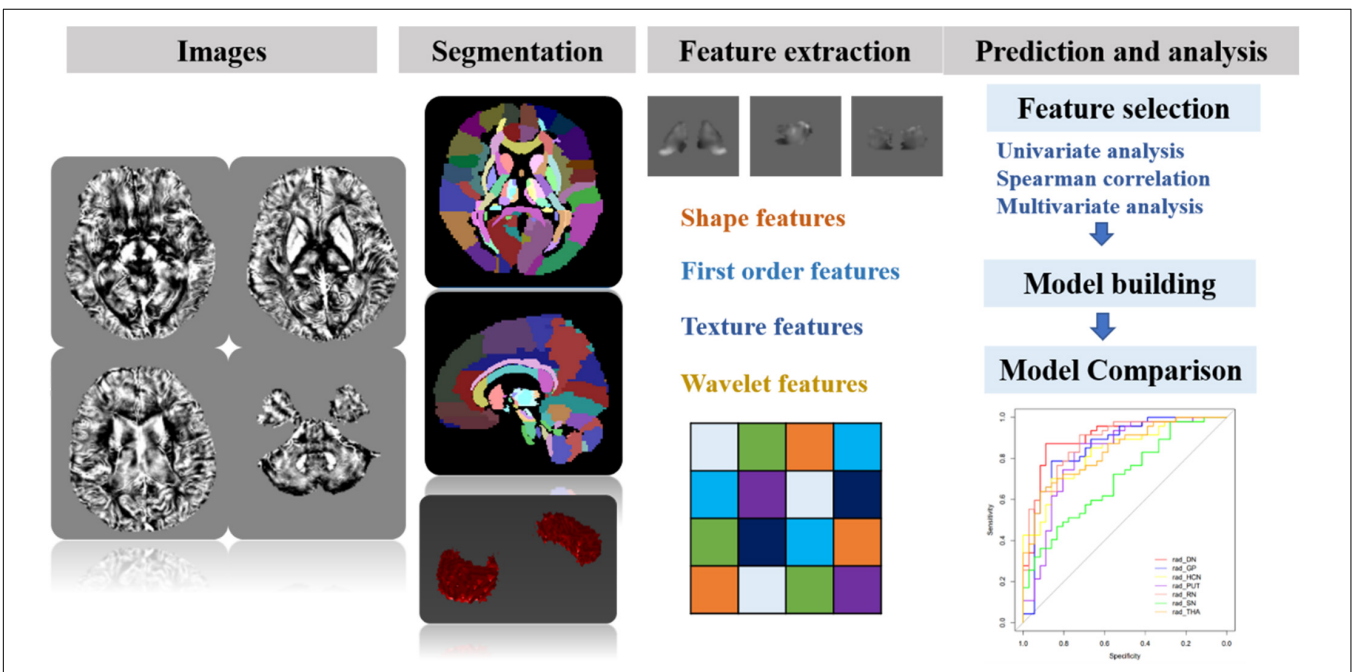


FIGURE 3 | The schematic flowchart of processing steps. Step 1: QSM reconstructions were performed. Step 2: after QSM atlas registration, seven ROIs were segmented on QSM images. Step 3: four sets of most representative features were chosen for generalizing and optimizing the model. Step 4: after feature selection, the prediction model for differentiating MS and NMOSD was built by combining radiomic features and clinical information. Furthermore, the classification performance was assessed with AUC using fivefold cross-validation.

Model Establishment and Performance Evaluation

The prediction models for differentiating MS and NMOSD were built with the logistic regression by the radiomic features only, demographic information (patient age, sex, disease duration, and EDSS score) and combining both above, respectively.

The area under the curves (AUC), sensitivity, specificity, and accuracy were used to evaluate the classification performance of the models. The classification performance was assessed with AUC using fivefold cross-validation. The DeLong test was used to compare the different radiomics models.

Statistical Analysis

Statistical analysis was performed by using the in-house software IPMs (IPM Statistics, version 2.4.0, GE Healthcare). Discrete data encoding the sexes of the patients were analyzed using *Pearson's chi-square* test. Two-sided two-sample *t*-tests or *Mann–Whitney U* tests were used to assess between-group differences for continuous demographic or clinical data, depending on whether they were normally distributed using the *Lilliefors* test. A two-tailed p-value < 0.05 represents the significant difference.

RESULTS

Demographic Data Analysis

The demographic and clinical characteristics of the enrolled patients are summarized in **Table 1**. Demographic information was collected at the time of visit including patient age, sex, disease duration, and EDSS score.

The NMOSD patients showed a greater female predominance, and higher EDSS scores, than the MS patients.

Radiomic Feature-Only Selection and Model Building

In this study, 874 radiomic features were extracted from the ROIs. First, with univariate analysis, the seven DGM regions (HCN, DN, GP, PUT, RN, SN, and THA) remained 51, 59, 206, 49, 78, 9, and 192 features, respectively. Next, the number of the 19, 21, 57, 27, 33, 7, and 67 features were then selected with the *Spearman* correlation analysis. Finally, with the multivariate analysis, the radiomic features were reduced to several most predictive features (HCN: 4, DN: 4, GP: 3, PUT: 3, RN: 5, SN: 2, THA: 3) in each region. Moreover, the selected radiomic features in different regions are shown in **Table 2**. The machine learning models with logistic regression were established with fivefold cross-validation.

TABLE 1 | The demographic and clinical data of participants.

Characteristics	MS	NMOSD	*p*-value
No. of patients	47	36	
Sex (male/female)	18/29	6/30	0.031[a]
Age (years)	40.00 ± 13.72	42.14 ± 12.34	0.464[b]
Disease duration (years)	5.42 (2.07, 9.19)	3.88 (1.29, 9.06)	0.301[c]
EDSS	2.00 (1.00, 3.00)	2.50 (2.00, 4.78)	0.002[c]

EDSS, Expanded Disability Status Scale; MS, multiple sclerosis; NMOSD, neuromyelitis optica spectrum disorder.
[a]*p-values obtained using the Pearson's chi-square test.*
[b]*p-values obtained using two-sample two-tailed t-tests.*
[c]*p-values obtained using two-tailed Mann–Whitney U-test.*

Model Evaluation

The performance of the prediction models assessed with the AUC, sensitivity, specificity, and accuracy were calculated. The DGM regions except for SN (AUC: 0.702, 95% CI: 0.603–0.793) had a better discrimination performance than the demographic data-only model and were more than 0.80. The model comparisons with the DeLong test were performed, as shown in **Table 3**; the results showed a significant difference that the p-value was 0.002 between DN and SN, 0.016 between GP and SN, and 0.004 between RN and SN, and there were no significant differences between other models. The DN reflected the best performance in all DGM regions with the AUC of 0.902 (95% CI: 0.840–0.955), a sensitivity of 0.851, a specificity of 0.889, and an accuracy of 0.867. All radiomics-only models' discrimination performance including AUC, sensitivity, specificity, and accuracy are shown in **Table 4**. The receiver operating characteristic curve (ROC) of all DGM regions is shown in **Figure 4**.

The demographic information-only model was built and showed an AUC of 0.733 (95% CI: 0.639–0.818), a sensitivity of 0.511, a specificity of 0.861, and an accuracy of 0.663. We used QSM-derived radiomic features, plus four demographic variables: sex, age, disease duration, and EDSS score to build a regression model to differentiate MS from NMOSD. The combined model showed a greater discrimination performance than the models mentioned above with an AUC of 0.927 (95% CI: 0.871–0.984). The ROC of the demographic information-only model, radiomics-only model of DN, and combined model are shown in **Figure 5**.

DISCUSSION

In the present study, we investigated the great performances of QSM-derived radiomic features in discriminating MS from NMOSD. Combining the radiomic features and demographic information to establish the hybrid model, it showed a higher discrimination performance, which can help the clinical differential diagnosis of these two diseases.

The diagnostic criteria of MS and NMOSD are constantly updated and recognized by the public, but MS showed inflammation, axonal loss, oligodendrocyte damage, gliosis, macrophage infiltration, microglia, and neurodegeneration, while NMOSD presented a perivascular immunoglobulin deposition and complement activation. Thus, it is still challenging to differentiate and diagnose these two diseases in clinical practice. Radiomics methods can provide us with more information. In previous studies, radiomics methods were based on traditional MRI techniques to differentiate MS from NMOSD. Liu et al. (2019) extracted radiomic features from the spinal cord lesions on sagittal T_2WI, and Ma et al. (2019) extracted the radiomics from lesions on T_2WI to differentiate

TABLE 2 | The selected radiomic features with different regions.

ROIs	Features
DN	original_glrlm_GrayLevelNonUniformity wavelet–LHH_glszm_GrayLevelVariance wavelet–HLL_glszm_GrayLevelVariance log–sigma-1–0-mm–3D_glszm_GrayLevelNonUniformity
GP	wavelet–LLH_glrlm_ShortRunHighGrayLevelEmphasis wavelet–LHL_gldm_DependenceEntropy wavelet–HLL_glszm_ZoneVariance log–sigma–1–0–mm–3D_glszm_ZoneEntropy
HCN	wavelet–HLH_glcm_JointAverage wavelet–HHL_glszm_HighGrayLevelZoneEmphasis wavelet–LLH_gldm_SmallDependenceHighGrayLevelEmphasis
PUT	wavelet–HHL_glrlm_ShortRunHighGrayLevelEmphasis wavelet–HHL_glcm_Imc1 wavelet–LLH_glszm_SizeZoneNonUniformityNormalized
RN	wavelet–LHH_gldm_SmallDependenceHighGrayLevelEmphasis wavelet–HLL_firstorder_Kurtosis wavelet–HLL_gldm_SmallDependenceHighGrayLevelEmphasis wavelet–HHL_glrlm_HighGrayLevelRunEmphasis wavelet–HLL_glszm_GrayLevelNonUniformity
SN	wavelet–LLH_glrlm_ShortRunHighGrayLevelEmphasis wavelet–HLL_firstorder_Skewness
THA	wavelet–LHL_glrlm_HighGrayLevelRunEmphasis wavelet–LLL_glrlm_ShortRunHighGrayLevelEmphasis wavelet–HHL_firstorder_Range

DN, dentate nucleus; GP, globus pallidus; HCN, head of the caudate nucleus; PUT, putamen; RN, red nucleus; SN, substantia nigra; THA, thalamus.

TABLE 3 | The DeLong test for the model comparisons.

Model	DeLong test					
	GP	HCN	PUT	RN	SN	THA
DN	0.434	0.187	0.216	0.754	0.0018*	0.289
GP		0.662	0.599	0.596	0.016*	0.727
HCN			0.877	0.312	0.083	0.891
PUT				0.289	0.107	0.743
RN					0.0042*	0.406
SN						0.042

**Represents significant comparison differences between models.*

TABLE 4 | Comparison of the different radiomics models.

Model	AUC (95% CI)	Sensitivity	Specificity	Accuracy
DN	0.902 (0.84, 0.955)	0.851	0.889	0.867
GP	0.856 (0.773, 0.928)	0.766	0.861	0.807
HCN	0.830 (0.759, 0.899)	0.681	0.861	0.759
PUT	0.821 (0.737, 0.901)	0.851	0.722	0.795
RN	0.885 (0.823, 0.941)	0.894	0.722	0.819
SN	0.702 (0.603, 0.793)	0.447	0.833	0.614
THA	0.838 (0.763, 0.906)	0.617	0.917	0.747

AUC, area under the receiver operating characteristic curve.

MS from NMOSD. There are different levels of iron in the brain. Therefore, the traditional MRI scans are based on the demyelination and inflammation of the lesions, but the iron deposition is also the pathological change of MS and NMOSD.

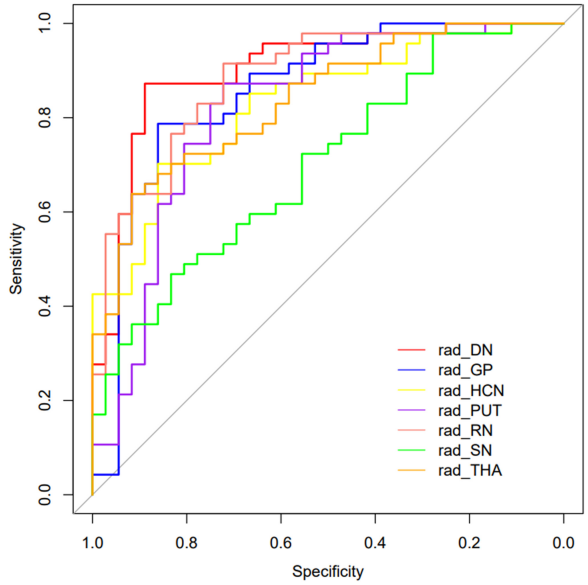


FIGURE 4 | The curves of the different model were shown. Different colors represented the model of different DGM regions.

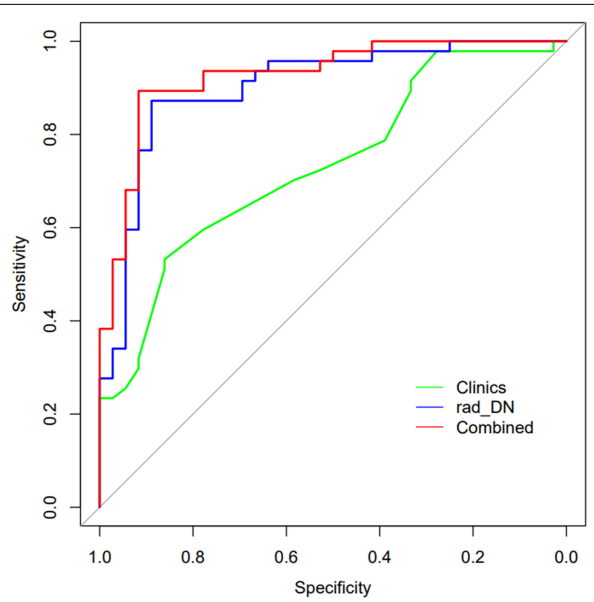


FIGURE 5 | The ROCs of the demographic information-only model, radiomics-only model of DN, and the combined model are shown.

Therefore, iron deposition is also essential in the differential diagnosis of these two diseases. Iron is an essential trace element for the human body which keeps many enzymes active, such as neurotransmitter synthesis, myelin formation, and mitochondrial energy metabolism. Its homeostasis is crucial for the normal physiological function of the brain. However, excess free iron catalyzes the generation of a large number of free radicals, which enhances oxygen toxicity and leads to

oxidative hyperactivity, thus participating in the pathogenesis of MS and NMOSD (Doring et al., 2016; Carocci et al., 2018; Zivadinov et al., 2018). As an advanced MRI technique for quantitative iron deposition, QSM-related studies (Chen et al., 2012; Langkammer et al., 2013; Rudko et al., 2014; Cobzas et al., 2015; Doring et al., 2016; Elkady et al., 2017; Hagemeier et al., 2017; Zivadinov et al., 2018; Pudlac et al., 2020) have also studied and confirmed the differences between the two diseases. Therefore, the differences in iron deposition are highly likely to change the image characteristics of QSM.

Previous studies have not reported the differences in quantitative susceptibility between the two diseases. To our knowledge, this is the first study that provides a comprehensive difference quantification of MS and NMOSD using QSM-derived radiomic features of the DGM. In our study, we extracted the most important radiomic features to establish the radiomics-only models. Furthermore, after fivefold cross-validation, we found that the DGM regions except for SN can differentiate the two diseases accurately with the AUC superior to 0.80. It indicated that the QSM-derived radiomic features do have the differences between MS and NMOSD patients and provide complementary information for clinical differential diagnosis of these two diseases.

In addition, the radiomics-only models based on the DN had the highest discrimination performance with an AUC of 0.902, using four radiomic features after feature selection. The gray-level run length matrix (GLRLM) gives the size of homogeneous runs in four directions for each gray level, and the gray-level zone length matrix (GLZLM) provides information on the size of homogeneous zones for each gray level in two dimensions. As the largest nucleus in the cerebellum, DN is located deep in the white matter of the cerebellum and has the functions of regulating body balance, regulating muscle tension, and coordinating voluntary movement. In MS patients, evidence has been found that the brain iron level is abnormal because iron accumulates in the basal ganglia area. Thus, we speculated that MS and NMOSD patients might have significant differences in athletic performance. Although the radiomic feature differences have been shown by DN in QSM images, there is no evidence that the two diseases have significant differences in the values of iron deposition in this area, which may be due to the long-term accumulation of iron deposition resulting in the changes in the images. However, due to the similarity of damage between the two diseases, the differences in quantitative susceptibility were not shown in previous studies.

The demographic information including patient age, sex, disease duration, and EDSS score was also collected in this work. The NMOSD patients showed a greater female predominance and higher EDSS scores than the MS patients. The demographic information-only model showed a lower AUC of 0.733 than any radiomics-only models in DGM regions except for SN. Combining the demographic information and QSM-derived radiomic features, we built a hybrid model that improved the discrimination performance and achieved the highest AUC of 0.927, highlighting the importance of comprehensive consideration of clinical and imaging features. Future work needs to clarify the underlying pathophysiological mechanism of the difference in QSM radiomic features between these two diseases, verify the effectiveness of this model, and promote the application of this model in clinical practice.

Several limitations also exist in this work. First, although we have extracted highly correlated radiomic features and obtained a relatively accurate discrimination performance, the pathophysiological mechanism is unclear. Second, our study only adopted QSM images reflecting iron deposition characteristics but did not include other sequences. Future work can use more sequences to obtain more information to identify MS and NMOSD. Finally, the relatively small sample size would affect the results of the discrimination model.

CONCLUSION

In conclusion, we have investigated the performance of QSM-derived radiomic features of the DGM to distinguish MS from NMOSD. The results of our study indicate that the discrimination model with the QSM-derived radiomic features in combination with demographic information has the potential to classify MS and NMOSD.

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 Institutional Review Board of the First Affiliated Hospital of Chongqing Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZY offered the researches idea and prepared the manuscript. HL offered the data processing support, which platforms and resources were assisted by the company. XC assisted in imaging processing. YZ assisted us in completing the statistics and experimental design. QZ, CZ, SD, YP, and YL provided guidance and critical reviews. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Medicine Scientific Key Research Project of Chongqing Health Bureau Grant (2016ZDXM002).

ACKNOWLEDGMENTS

We would like to thank all the subjects who participated in this study.

REFERENCES

Calabrese, M., Oh, M. S., Favaretto, A., Rinaldi, F., Poretto, V., Alessio, S., et al. (2012). No MRI evidence of cortical lesions in neuromyelitis optica. *Neurology* 79, 1671–1676. doi: 10.1212/WNL.0b013e31826e9a96

Carocci, A., Catalano, A., Sinicropi, M. S., and Genchi, G. (2018). Oxidative stress and neurodegeneration: the involvement of iron. *Biometals* 31, 715–735. doi: 10.1007/s10534-018-0126-2

Chen, X., Zeng, C., Luo, T., Ouyang, Y., Lv, F., Rumzan, R., et al. (2012). Iron deposition of the deep grey matter in patients with multiple sclerosis and neuromyelitis optica: a control quantitative study by 3D-enhanced susceptibility-weighted angiography (ESWAN). *Eur. J. Radiol.* 81, e633–e639. doi: 10.1016/j.ejrad.2012.01.003

Cobzas, D., Sun, H., Walsh, A. J., Lebel, R. M., Blevins, G., and Wilman, A. H. (2015). Subcortical gray matter segmentation and voxel-based analysis using transverse relaxation and quantitative susceptibility mapping with application to multiple sclerosis. *J. Magn. Reson. Imaging* 42, 1601–1610. doi: 10.1002/jmri.24951

Doring, T. M., Granado, V., Rueda, F., Deistung, A., Reichenbach, J. R., Tukamoto, G., et al. (2016). Quantitative susceptibility mapping indicates a disturbed brain iron homeostasis in neuromyelitis optica - a pilot study. *PLoS One* 11:e0155027. doi: 10.1371/journal.pone.0155027

Elkady, A. M., Cobzas, D., Sun, H., Blevins, G., and Wilman, A. H. (2017). Progressive iron accumulation across multiple sclerosis phenotypes revealed by sparse classification of deep gray matter. *J. Magn. Reson. Imaging* 46, 1464–1473. doi: 10.1002/jmri.25682

Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: images Are more than pictures. They are data. *Radiology* 278, 563–577. doi: 10.1148/radiol.2015151169

Haacke, E. M., Cheng, N. Y., House, M. J., Liu, Q., Neelavalli, J., Ogg, R. J., et al. (2005). Imaging iron stores in the brain using magnetic resonance imaging. *Magn. Reson. Imaging* 23, 1–25. doi: 10.1016/j.mri.2004.10.001

Hagemeier, J., Zivadinov, R., Dwyer, M. G., Polak, P., Bergsland, N., Weinstock-Guttman, B., et al. (2017). Changes of deep gray matter magnetic susceptibility over 2 years in multiple sclerosis and healthy control brain. *Neuroimage Clin.* 18, 1007–1016. doi: 10.1016/j.nicl.2017.04.008

Huh, S. Y., Min, J. H., Kim, W., Kim, S. H., Kim, H. J., Kim, B. J., et al. (2014). The usefulness of brain MRI at onset in the differentiation of multiple sclerosis and seropositive neuromyelitis optica spectrum disorders. *Mult. Scler.* 20, 695–704. doi: 10.1177/1352458513506953

Jarius, S., and Wildemann, B. (2013). Aquaporin-4 antibodies (NMO-IgG) as a serological marker of neuromyelitis optica: a critical review of the literature. *Brain Pathol.* 23, 661–683. doi: 10.1111/bpa.12084

Jurynczyk, M., Craner, M., and Palace, J. (2015). Overlapping CNS inflammatory diseases: differentiating features of NMO and MS. *J. Neurol. Neurosurg. Psychiatry* 86, 20–25. doi: 10.1136/jnnp-2014-308984

Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., et al. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48, 441–446. doi: 10.1016/j.ejca.2011.11.036

Langkammer, C., Liu, T., Khalil, M., Enzinger, C., Jehna, M., Fuchs, S., et al. (2013). Quantitative susceptibility mapping in multiple sclerosis. *Radiology* 267, 551–559. doi: 10.1148/radiol.12120707

Lennon, V. A., Wingerchuk, D. M., Kryzer, T. J., Pittock, S. J., Lucchinetti, C. F., Fujihara, K., et al. (2004). A serum autoantibody marker of neuromyelitis optica: distinction from multiple sclerosis. *Lancet* 364, 2106–2112. doi: 10.1016/S0140-6736(04)17551-X

Liu, Y., Dong, D., Zhang, L., Zang, Y., Duan, Y., Qiu, X., et al. (2019). Radiomics in multiple sclerosis and neuromyelitis optica spectrum disorder. *Eur. Radiol.* 29, 4670–4677. doi: 10.1007/s00330-019-06026-w

Liu, Y., Duan, Y., Huang, J., Ren, Z., Ye, J., Dong, H., et al. (2015). Multimodal quantitative mr imaging of the thalamus in multiple sclerosis and neuromyelitis optica. *Radiology* 277, 784–792. doi: 10.1148/radiol.2015142786

Ma, X., Zhang, L., Huang, D., Lyu, J., Fang, M., Hu, J., et al. (2019). Quantitative radiomic biomarkers for discrimination between neuromyelitis optica spectrum disorder and multiple sclerosis. *J. Magn. Reson. Imaging* 49, 1113–1121. doi: 10.1002/jmri.26287

Noseworthy, J. H., Lucchinetti, C., Rodriguez, M., and Weinshenker, B. G. (2000). Multiple sclerosis. *N. Engl. J. Med.* 343, 938–952. doi: 10.1056/NEJM200009283431307

Palace, J., Leite, M. I., Nairne, A., and Vincent, A. (2010). Interferon beta treatment in neuromyelitis optica: increase in relapses and aquaporin 4 antibody titers. *Arch. Neurol.* 67, 1016–1017. doi: 10.1001/archneurol.2010.188

Pudlac, A., Burgetova, A., Dusek, P., Nytrova, P., Vaneckova, M., Horakova, D., et al. (2020). Deep gray matter iron content in neuromyelitis optica and multiple sclerosis. *Biomed. Res. Int.* 2020:6492786. doi: 10.1155/2020/6492786

Rudko, D. A., Solovey, I., Gati, J. S., Kremenchutzky, M., and Menon, R. S. (2014). Multiple sclerosis: improved identification of disease-relevant changes in gray and white matter by using susceptibility-based MR imaging. *Radiology* 272, 851–864. doi: 10.1148/radiol.14132475

Schweser, F., Deistung, A., Lehr, B. W., and Reichenbach, J. R. (2011). Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: an approach to in vivo brain iron metabolism? *Neuroimage* 54, 2789–2807. doi: 10.1016/j.neuroimage.2010.10.070

Sinnecker, T., Dörr, J., Pfueller, C. F., Harms, L., Ruprecht, K., Jarius, S., et al. (2012). Distinct lesion morphology at 7-T MRI differentiates neuromyelitis optica from multiple sclerosis. *Neurology* 79, 708–714. doi: 10.1212/WNL.0b013e3182648bc8

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., et al. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 17, 162–173. doi: 10.1016/S1474-4422(17)30470-2

Wingerchuk, D. M., Banwell, B., Bennett, J. L., Cabre, P., Carroll, W., Chitnis, T., et al. (2015). International panel for NMO diagnosis. International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. *Neurology* 85, 177–189. doi: 10.1212/WNL.0000000000001729

Wingerchuk, D. M., Lennon, V. A., Lucchinetti, C. F., Pittock, S. J., and Weinshenker, B. G. (2007). The spectrum of neuromyelitis optica. *Lancet Neurol.* 6, 805–815. doi: 10.1016/S1474-4422(07)70216-8

Yip, S. S., and Aerts, H. J. (2016). Applications and limitations of radiomics. *Phys. Med. Biol.* 61, R150–R166. doi: 10.1088/0031-9155/61/13/R150

Zhang, Y., Wei, H., Cronin, M. J., He, N., Yan, F., and Liu, C. (2018). Longitudinal atlas for normative human brain development and aging over the lifespan using quantitative susceptibility mapping. *Neuroimage* 171, 176–189. doi: 10.1016/j.neuroimage.2018.01.008

Zivadinov, R., Tavazzi, E., Bergsland, N., Hagemeier, J., Lin, F., Dwyer, M. G., et al. (2018). Brain iron at quantitative MRI is associated with disability in multiple sclerosis. *Radiology* 289, 487–496. doi: 10.1148/radiol.2018180136

Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J. W. L., Andrearczyk, V., Apte, A., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295, 328–338. doi: 10.1148/radiol.2020191145

Conflict of Interest: HL was employed by the company GE Healthcare, China.

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 08 December 2021
doi: 10.3389/fnins.2021.796172

Frontal Cortex Segmentation of Brain PET Imaging Using Deep Neural Networks

Qianyi Zhan[1,2], *Yuanyuan Liu*[1,2], *Yuan Liu*[1,2] *and Wei Hu*[3]*

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China, [2] Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi, China, [3] Department of Nuclear Medicine, Nanjing Medical University, Affiliated Wuxi People's Hospital, Wuxi, China

^{18}F-FDG positron emission tomography (PET) imaging of brain glucose use and amyloid accumulation is a research criteria for Alzheimer's disease (AD) diagnosis. Several PET studies have shown widespread metabolic deficits in the frontal cortex for AD patients. Therefore, studying frontal cortex changes is of great importance for AD research. This paper aims to segment frontal cortex from brain PET imaging using deep neural networks. The learning framework called **F**rontal cortex **S**egmentation model of brain **PET** imaging (FSPET) is proposed to tackle this problem. It combines the anatomical prior to frontal cortex into the segmentation model, which is based on conditional generative adversarial network and convolutional auto-encoder. The FSPET method is evaluated on a dataset of 30 brain PET imaging with ground truth annotated by a radiologist. Results that outperform other baselines demonstrate the effectiveness of the FSPET framework.

Keywords: brain image segmentation, convolutional auto-encoder, conditional generative adversarial network, PET, Alzheimer's disease

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Yanhui Zhang,
Hebei University of Chinese Medicine, China
Qiusheng Shen,
Changshu Affiliated Hospital of Nanjing University of Traditional Chinese Medicine, China

***Correspondence:**
Wei Hu
huwei@njmu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *16 October 2021*
Accepted: *01 November 2021*
Published: *08 December 2021*

Citation:
Zhan Q, Liu Y, Liu Y and Hu W (2021) Frontal Cortex Segmentation of Brain PET Imaging Using Deep Neural Networks. Front. Neurosci. 15:796172. doi: 10.3389/fnins.2021.796172

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive disease that destroys memory and other important mental functions. As of 2019, it ranked as the sixth leading cause of death in China (Vos et al., 2020). There are more than 10 million patients with AD in China, a country with the most AD patients in the world (Jia et al., 2020).

AD is usually diagnosed based on the clinical manifestation. Nowadays, medical imaging including computed tomography (CT) or magnetic resonance imaging (MRI), and with single-photon emission computed tomography (SPECT) or positron emission tomography (PET), can be used to help doctors understand the pathophysiology of AD, for example, Aβ plaques, neurofibrillary tangles, and neuroinflammation. Moreover, the pathophysiology of AD is believed that starts years ahead of the of clinical observation, and helps detect AD earlier than conventional diagnostic tools (Marcus et al., 2014).

Among the above medical imaging technique, PET/CT is a nuclear medicine technique that combines a PET scanner and a CT scanner to acquire sequential images from both devices in the same session, which are combined into a single superposed image. **Figure 1** shows the brain PET/CT fusion image. The first line is the PET imaging, and the second line is the CT imaging. The fusion imaging of PET/CT is list in the third line. Each line from left to right is (a) coronal section, (b) median sagittal section, and (c) transverse section.

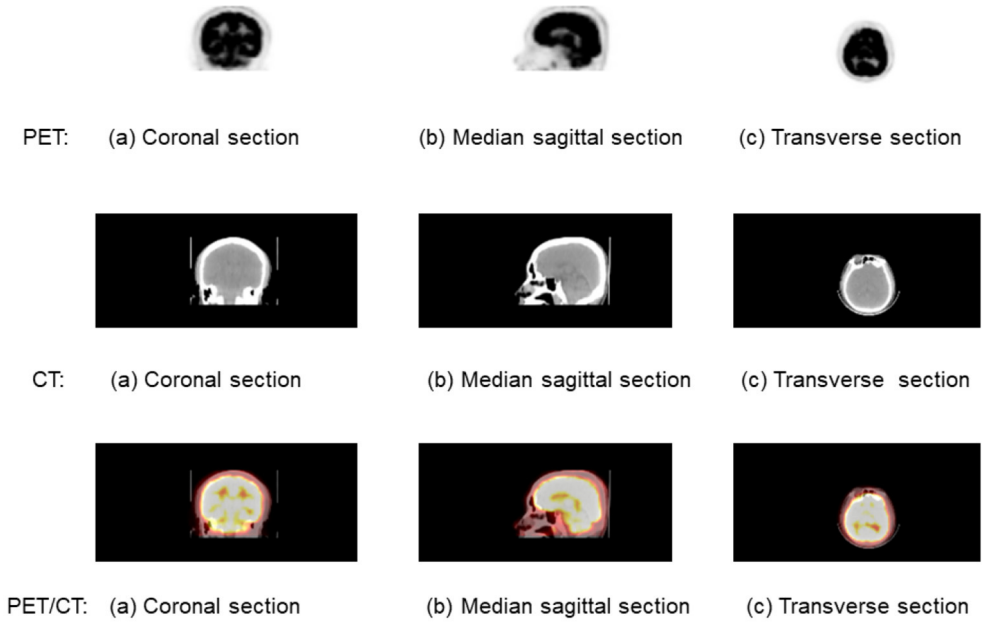


FIGURE 1 | A brain positron emission tomography (PET)/computed tomography (CT) fusion image. The first line is the PET imaging, and the second line is the CT imaging. The fusion imaging of PET/CT is list in the third line. Each line from left to right is (a) coronal section, (b) median sagittal scan, and (c) transverse section.

^{18}F-FDG PET imaging of brain glucose use and amyloid accumulation is a research criteria for AD diagnosis (Berti et al., 2011). Several ^{18}F-FDG PET studies have been conducted to estimate AD-related brain changes. They have consistently shown widespread metabolic deficits in the neocortical association areas, such as frontal cortex. Further studies have demonstrated that CMRglc in frontal cortex suffers an average decline of 16 − 19% over a 3-year period (Smith et al., 1992; Mielke et al., 1994). Frontal cortex covers frontal lobe and contains most of the dopamine neurons. **Figure 2** shows the location of frontal cortex in the brain. The yellow part of the left subfigure is its anatomical location, while the red contour in the right subfigure indicates its location in ^{18}F-FDG PET imaging. Due to its sensitive to detect frontal cortex changes over time, ^{18}F-FDG PET imaging can be used not only for AD diagnosis but also to monitor dementia progression and therapeutic interventions. Therefore, PET imaging are valuable in the assessment of patients with AD. Moreover, the frontal cortex segmentation of PET imaging is crucial for understanding AD progression on AD-related regions in brain.

Although frontal cortex segmentation is an important problem for AD research. However as far as we know, this paper is the first work that studies the frontal cortex segmentation problem for PET imaging. Unlike organ or tumor, which is different from other tissue with gray-level, texture, gradients, edges, shape, etc., frontal cortex is a part of brain without obvious boundaries. Moreover, supervised learning frameworks need segmentation ground truth from professional doctor, and it is difficult to get large number of annotated imaging. All these makes frontal cortex segmentation a tough problem.

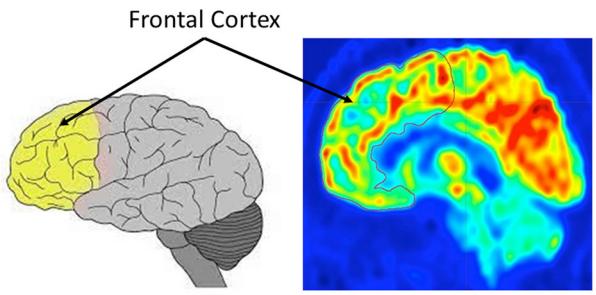


FIGURE 2 | The frontal cortex in the brain: the left is anatomical location, and the right is for ^{18}F-FDG positron emission tomography (PET) imaging.

Since manual segmentation is time consuming, automatic semantic segmentation for medical images, which makes pathological structures changes clear in images, becomes one of the hottest research topic in image processing. Currently, more and more machine learning technologies have been used in medical applications, such as medical single processing, medical image processing, medical data analyzing, and so on (Jiang et al., 2021a,b; Yang et al., 2021). Brain and brain tumor segmentation is one of the most popular medical image segmentation tasks (Szilagyi et al., 2003; Tu and Bai, 2009; Zhang et al., 2015; Jiang et al., 2019). Many approaches have been proposed to address this problem, such as thresholding (Sujji et al., 2013), edge detection (Tang et al., 2000), Markov random fields

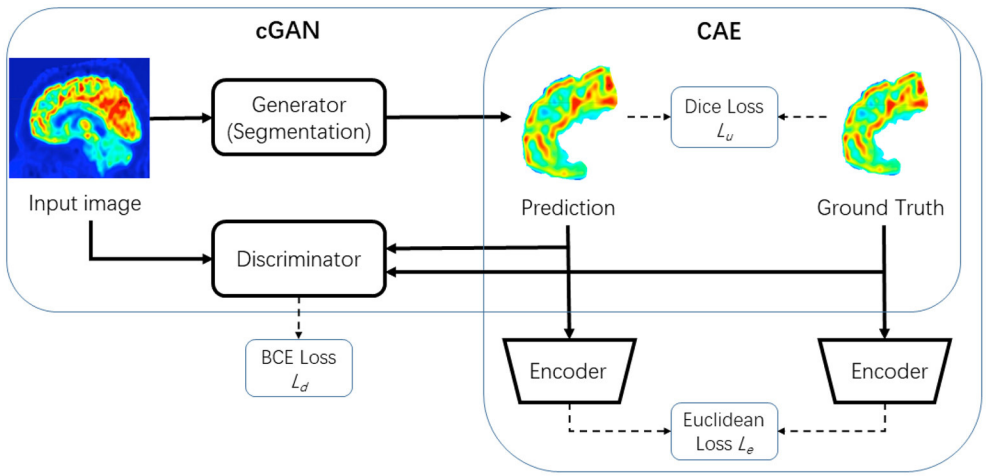


FIGURE 3 | Framework of proposed model FSPET based on conditional generative adversarial network (cGAN) and convolutional auto-encoder (CAE).

(MRF) (Held et al., 1997), and support vector machine (SVM) (Akselrod-Ballin et al., 2006).

Due to the rapid development of deep learning, neural networks, which can extract hierarchical feature of images, become one of the most effective technique in brain imaging segmentation (Fakhry et al., 2016; Işın et al., 2016; Zhao et al., 2018). U-net (Ronneberger et al., 2015) and its 3D version V-Net (Milletari et al., 2016) are the most well-known deep learning architecture in medical image segmentation. Recently, organ and tissue's shape and position priors are combined into the segmentation algorithm to improve the accuracy. (Oktay et al., 2017) proposes a training framework ACNN, which incorporates cardiac anatomical prior into CNN. Boutillon et al. (2019) combines scapula bone anatomical prior into a conditional adversarial learning method.

The related work has made large progress in semantic segmentation in medical imaging. However, they are not designed for frontal cortex segmentation in PET imaging, and they cannot be utilized directly for this problem. Motivated by this, in this paper, we propose the supervised segmentation framework: **F**rontal cortex **S**egmentation model of brain **PET** imaging (FSPET). The FSPET model based on both conditional generative adversarial network (cGAN) and convolutional auto-encoder (CAE) incorporates the anatomical prior to improve the prediction accuracy.

The contribution of FSPET dedicated to frontal cortex segmentation is threefold. First, the CAE is used to find the embedding of frontal cortex shape priors in latent space. Second, the segmentation method based on U-net, as the generator of cGAN, learns the feature of frontal cortex to generate the binary mask in PET imaging. Third, the anatomical prior is fused into the discriminator model in cGAN to get more precise prediction. Extensive experiments demonstrate the effectiveness of the proposed FSPET model.

2. METHODS

In this section, we will introduce the proposed FSPET framework in detail, which combines the prior of frontal cortex shape in the deep neural networks, as shown in **Figure 3**. FSPET contains two parts: cGAN and CAE.

2.1. Conditional Generative Adversarial Networks

Generative Adversarial Network (GAN) (Goodfellow et al., 2014) is widely used for data augmentation by generating new images. Since PET imaging shows low contrast, low resolution, and blurred boundaries between different tissues, GAN is becoming a popular method for medical image segmentation (Luc et al., 2016; Son et al., 2017; Souly et al., 2017).

cGAN (Mirza and Osindero, 2014) is an extension of GAN, which is used as a machine learning framework for training generative models. The proposed FSPET model adopts the framework in Conze et al. (2021) based on cGAN, which consists of two neural networks: the generator G and the discriminator D.

The generator G of cGAN in the FSPET model is the segmentation framework, which learns the feature of frontal cortex to generate the binary mask in PET imaging. Formally, let x be the source image and y be the ground truth image of class label $y_i \in \mathcal{L} = \{1, 2, \ldots, c\}$. The generator learns the mapping between images and labels $G : x \rightarrow \mathcal{L}$ by optimizing the loss function using stochastic gradient descent. The generator of cGAN is often based on U-net framework. The network consists of a contracting path and an expansive path (shown in **Figure 4B**). The contracting path is a convolutional network that consists of repeated application of 3×3 convolutions, each followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation. Dice loss is used in U-net to compare the prediction $G(x)$ and ground truth y, in which the loss function is as follows:

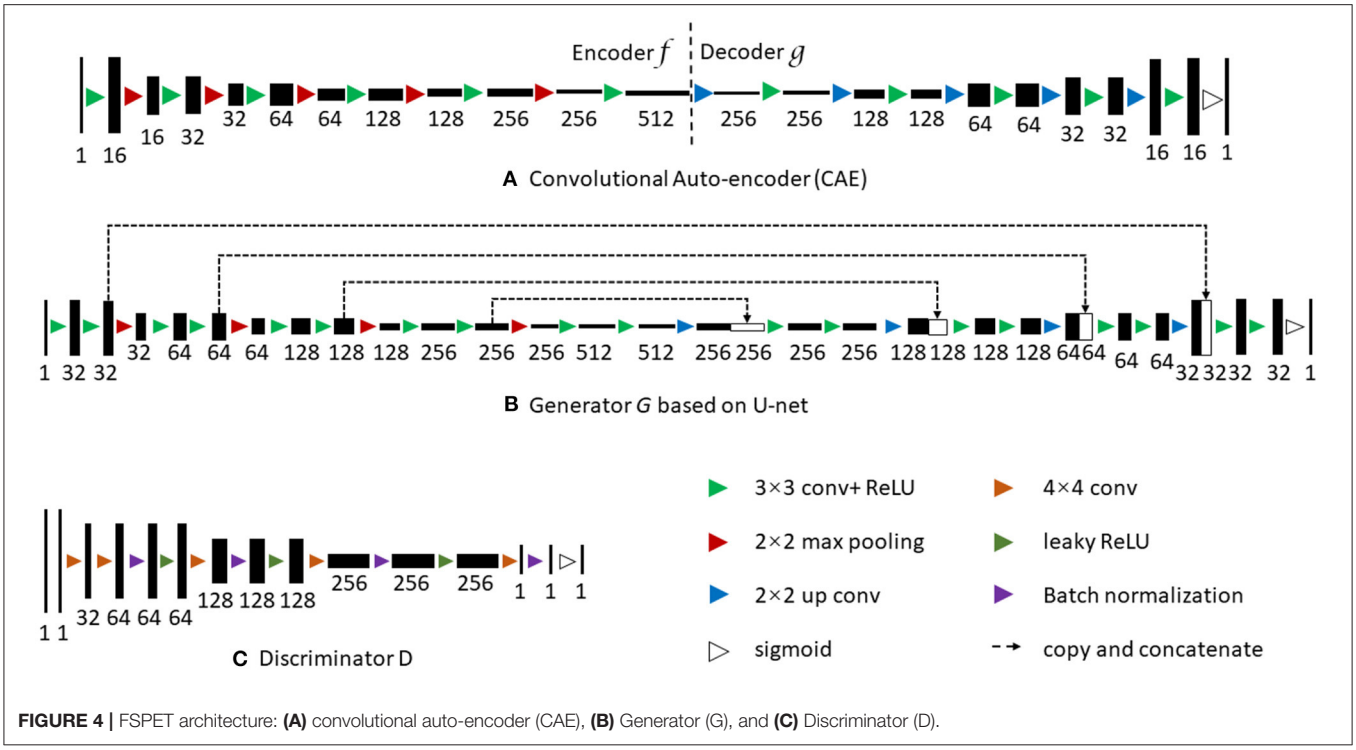


FIGURE 4 | FSPET architecture: **(A)** convolutional auto-encoder (CAE), **(B)** Generator (G), and **(C)** Discriminator (D).

TABLE 1 | Loss in the proposed FSPET model.

Notation	Loss type	Evaluated by	Calculated from	Loss function
L_u	Dice	Prediction and ground truth	U-net G	$\mathbb{E}_{x,y}[L_{dice}(G(x), y)]$
L_d	BCE	Prediction and input	Discriminator D	$\mathbb{E}_{x,y}[-log(D(x, G(x)))]$
L_e	Euclidean	Prediction and ground truth	encoder in CAE f	$\mathbb{E}_{x,y}[(f(G(x)) - f(y))^2]$

$$L_u = \mathbb{E}_{x,y}[L_{dice}(G(x), y)]. \quad (1)$$

The discriminator D in FSPET (shown in **Figure 4C**) inputs are source images and prediction to be evaluated. D distinguishes the given boundary by the generator from the realistic segmentation. The output is a binary prediction as to whether the image is real (class = 1) or fake (class = 0). In cGAN, binary cross entropy (BCE) loss is used to determine the loss function:

$$L_d = \mathbb{E}_{x,y}[-log(D(x, G(x)))]. \quad (2)$$

2.2. Convolutional Auto-Encoder

Auto-encoder is a type of neural networks used to learn a representation (encoding) for a set of data. It imposes a bottleneck in the network, which forces a compressed knowledge representation of the original input. An auto-encoder consists of two parts, the encoder and the decoder, which can be defined as f and g such that:

$$f : \mathcal{X} \rightarrow \mathbf{h}$$
$$g : \mathbf{h} \rightarrow \mathcal{X}$$
$$f, g = arg\min_{f,g} ||\mathcal{X} - (f \circ g)\mathcal{X}||,$$

where $\mathcal{X}$ is the input and $\mathbf{h}$ is usually referred to as code, the latent representation of the input. Motivated by Oktay et al. (2017), we utilized the CAE to find the embedding of frontal cortex shape priors in latent space (shown in **Figure 4A**). The BCE loss is minimized in the CAE framework with the ground truth y as input:

$$L_{CAE}(f, g) = \mathbb{E}_{x,y}[-log(y, g(f(y)))]. \quad (3)$$

As shown in **Figure 4A**, after CAE is fixed, we use its encoder part f for segmentation training. By conducting CAE low-dimensional projection on both prediction and ground truth, we can minimize the loss function as:

$$L_e = \mathbb{E}_{x,y}[(f(G(x)) - f(y))^2]. \quad (4)$$

2.3. Fusion

We have obtained three loss functions from different parts of the FSPET model respectively, whose information are listed in **Table 1**. Finally, we fuse the U-net segmentation method with frontal cortex shape priors. In the backward propagation, the loss

TABLE 2 | Quantitative assessment of U-net (Ronneberger et al., 2015), ACNN (Oktay et al., 2017), cGAN-Unet (Singh et al., 2018), and the FSPET model.

Model	Dice	Jaccard	Sensitive	Specificity	HD
U-net	71.03 ± 21.37	55.04 ± 20.43	72.29 ± 26.76	**98.22 ± 2.12**	38.73 ± 30.46
ACNN	74.57 ± 18.34	59.45 ± 19.17	78.37 ± 23.85	97.24 ± 1.88	35.48 ± 27.83
cGAN-Unet	79.04 ± 19.59	65.34 ± 14.45	79.75 ± 21.53	97.18 ± 2.57	30.32 ± 29.12
FSPET	**83.37 ± 15.98**	**71.47 ± 15.32**	**81.97 ± 20.88**	96.93 ± 2.27	**28.05 ± 25.78**

Bold results indicate the best scores.

function of generator G is:

$$\begin{aligned} L_G(G,D,f) = &\quad L_d + \lambda_1 L_u + \lambda_2 L_e \\ = &\quad \mathbb{E}_{x,y}[-log(D(x,G(x)))] \\ + &\quad \lambda_1 \mathbb{E}_{x,y}[L_{dice}(G(x),y)] \\ + &\quad \lambda_2 \mathbb{E}_{x,y}[(f(G(x)) - f(y))^2], \end{aligned} \tag{5}$$

where λ_1 and λ_2 are the weighting factor. Minimizing L_u tends to provide rough frontal cortex shape prediction, while maximizing $log(D(x, G(x)))$ is designed to improve contour delineations. At the same time, the latent loss L_e guarantees the global consistent and precise prediction similar to the original segmentation.

Additionally, the loss function of discriminator D is:

$$\begin{aligned} L_D(G,D) = &\quad \mathbb{E}_{x,y}[-log(D(x,y))] \\ + &\quad \mathbb{E}_{x,y}[-log(1 - D(x,G(x)))]. \end{aligned} \tag{6}$$

It maximizes $log(D(x, y)$, which is the loss between input and ground truth. Simultaneously, it minimizes loss value for generated $-log(1 - D(x, G(x)))$ masks. The optimization proceeds in alternative periods on G and D using stochastic gradient descent.

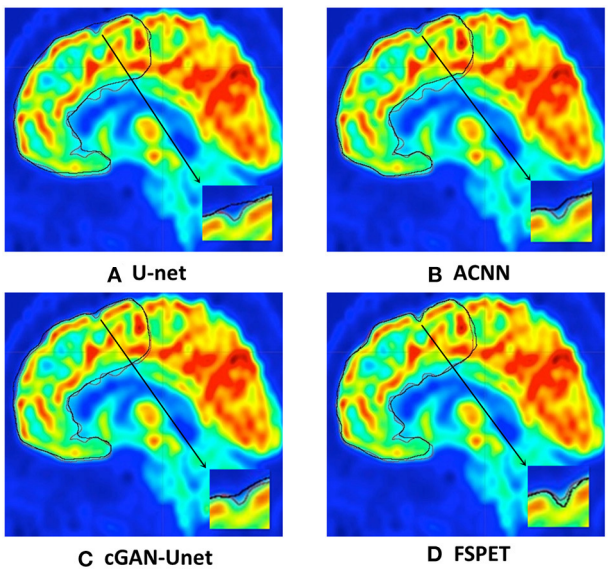


FIGURE 5 | Frontal cortex segmentation in median sagittal section of brain positron emission tomography (PET) imaging using U-net, ACNN, cGAN-Unet, and the FSPET model. Ground truth and predicted contour are in red and black, respectively. **(A)** U-net. **(B)** ACNN. **(C)** cGAN-Unet. **(D)** FSPET.

3. EXPERIMENTS

In this section, we conduct extensive experiments to validate the effectiveness of FSPET.

3.1. Validation Setup

Dataset: We collected 30 ^{18}F-FDG PET images from different patients and their sensitive information was erased. The brain PET images were acquired using a PET/CT (Discovery STE, General Electric, Waukesha, USA) approximately 1 h after an intravenous injection of ^{18}F-FDG (10 mCi). The original data were stored in Digital Imaging and Communications in Medicine (DICOM) format. Images were then resampled with a resolution 512 × 512 pixels. Frontal cortex in all images were annotated by a radiologist with 7 years experience to obtain the ground truth and also the shape priors.

Baselines: We compare FSPET with different baseline methods in frontal cortex segmentation for brain PET imaging. The comparison methods used in the experiments include:

- **U-net** (Ronneberger et al., 2015): U-net is the classical segmentation algorithm for medical images, and the generator G of FSPET is based on U-net. The architecture (shown in **Figure 4B**) contains contraction path and symmetric expanding path. The former path is used to capture the context in the image and the latter one is used to enable precise localization using transposed convolutions.
- **ACNN** (Oktay et al., 2017): It utilizes the auto-encoder and T-L network to combine anatomical prior knowledge into CNNs. These regularizers make predictions that are in agreement with the shape priors.
- **cGAN-Unet** (Singh et al., 2018): It is proposed for breast mass segmentation in mammography. The cGAN is used in the segmentation framework, in which the generative network learns the features of tumors and the adversarial network guarantees the contour to be similar to the ground truth.

Measurements: With the definition of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), the following metrics are used to provide an overall assessment of all methods:

- Dice coefficient (dice): It is a similarity measure over prediction and ground truth. It ranges between 0 and 1.

$$DSC = \frac{2TP}{2TP + FP + FN}$$

- Jaccard index (Jaccard): It is another similarity metric with range [0, 1].

$$J = \frac{TP}{TP + FP + FN}$$

- Sensitivity: It is a measure of how well a test can identify true positives.

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

- Specificity: It is a measure of how well a test can identify true negatives.

$$TNR = \frac{TN}{TN + FP}$$

- Hausdorff distance (HD): It is the greatest of all the distances from a point in one set to the closest point in the other set. HD measures how far two contours of prediction and ground truth are from each other. With A and B are the set of non-zero voxels in labels images, HD is defined as:

$$HD(X, Y) = max(h(A, B), h(B, A))$$
$$\text{where} \quad h(A, B) = max_{a \in A} min_{b \in B} ||a - b||$$

Training Detail: In this paper, we use Adam optimization (Kingma and Ba, 2014), which is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments. To train the FSPET model, the CAE with BCE loss is first optimized based on (3). With the learning rate of 0.01 for 30 epochs, the batch size is fixed at 32. Then the optimization of cGAN proceeds in alternative periods on G and D according to (5) and (6). In (5), the weight $\lambda_1 = 10^{-2}$ and $\lambda_2 = 10^{-4}$. With the learning rate of 10^{-4}, batch size 32 and 30 epochs enjoyed the best performance.

3.2. Results

Quantitative metric and score values are provided in **Table 2** for frontal cortex segmentation. When comparing U-net and ACNN, the dice score improves from 71.03% to 74.57%, and HD score decreases from 38.73 to 35.48. This demonstrates that extending U-net with CAE allows the model taking advantage of latent representation of shape priors. Moreover, significant improvements can be noticed using cGAN-Unet comparing with U-net (38.73 to 30.32 on HD), which indicates the appropriateness of embedding U-net into a cGAN pipeline. Combining CAE and cGAN networks, the proposed FSPET model discriminates more efficiently frontal cortex from surrounding structures by achieving the best score with regard to dice, Jaccard index, sensitivity, and HD. In particular, large gains in terms of Jaccard index (55.04–71.47%) and HD (38.73–28.05) are reported between U-net and FSPET.

Qualitative results for frontal cortex segmentation in median sagittal section of brain PET imaging are displayed in **Figure 5**. Compared to U-net, ACNN, and cGAN-Unet, which are prone to under- or over-segmentation, sometimes combined with unrealistic shapes, better contour adherence and shape consistency are reached by the FSPET model. We also take one example (shown in bottom right of each subfigure) for comparison, and the FSPET captures more complex shape and subtle contours compared to other frameworks. This reveals the importance of combining both adversarial networks and shape priors in the segmentation task.

4. CONCLUSION

This paper propose a deep learning framework to segment frontal cortex from brain PET imaging. The model based on both cGAN and CAE incorporates the anatomical prior to improve the prediction accuracy. Future work will utilize the proposed method to detect other parts of AD-related brain area, such as hippocampus.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Jiangnan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

QZ contributed to the conception of the study and wrote the manuscript. YuanyuanL performed the experiment. YuanL helped to perform the analysis with constructive discussions. WH contributed significantly to data preparation.

REFERENCES

Akselrod-Ballin, A., Galun, M., Gomori, M. J., Basri, R., and Brandt, A. (2006). "Atlas guided identification of brain structures by combining 3D segmentation and SVM classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Copenhagen: Springer), 209–216. doi: 10.1007/11866763_26

Berti, V., Pupi, A., and Mosconi, L. (2011). PET/CT in diagnosis of dementia. *Ann. N. Y. Acad. Sci.* 1228:81. doi: 10.1111/j.1749-6632.2011.06015.x

Boutillon, A., Borotikar, B., Burdin, V., and Conze, P. H. (2019). "Combining shape priors with conditional adversarial networks for improved scapula segmentation in MR images," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (Iowa City, IA). doi: 10.1109/ISBI45749.2020.9098360

Conze, P.-H., Kavur, A. E., Cornec-Le Gall, E., Gezer, N. S., Le Meur, Y., Selver, M. A., et al. (2021). Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks. *Artif. Intell. Med.* 117:102109. doi: 10.1016/j.artmed.2021.102109

Fakhry, A., Peng, H., and Ji, S. (2016). Deep models for brain EM image segmentation: novel insights and improved performance. *Bioinformatics* 32, 2352–2358. doi: 10.1093/bioinformatics/btw165

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 27.

Held, K., Kops, E. R., Krause, B. J., Wells, W. M., Kikinis, R., and Muller-Gartner, H.-W. (1997). Markov random field segmentation of brain MR images. *IEEE Trans. Med. Imaging* 16, 878–886. doi: 10.1109/42.650883

Işın, A., Direkoğlu, C., and Şah, M. (2016). Review of MRI-based brain tumor image segmentation using deep learning methods. *Proc. Comput. Sci.* 102, 317–324. doi: 10.1016/j.procs.2016.09.407

Jia, L., Du, Y., Chu, L., Zhang, Z., Li, F., Lyu, D., et al. (2020). Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. *Lancet Public Health* 5, e661–e671. doi: 10.1016/S2468-2667(20)30185-7

Jiang, Y., Gu, X., Wu, D., Hang, W., Xue, J., Qiu, S., et al. (2021a). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain ct image segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 40–52. doi: 10.1109/TCBB.2019.2963873

Jiang, Y., Zhang, Y., Lin, C., Wu, D., and Lin, C.-T. (2021b). EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system. *IEEE Trans. Intell. Transport. Syst.* 22, 1752–1764. doi: 10.1109/TITS.2020.2973673

Jiang, Y., Zhao, K., Xia, K., Xue, J., Zhou, L., Ding, Y., et al. (2019). A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation. *J. Med. Syst.* 43, 1–9. doi: 10.1007/s10916-019-1245-1

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Luc, P., Couprie, C., Chintala, S., and Verbeek, J. (2016). Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408*.

Marcus, C., Mena, E., and Subramaniam, R. M. (2014). Brain pet in the diagnosis of Alzheimer's disease. *Clin. Nuclear Med.* 39, 423–426. doi: 10.1097/RLU.0000000000000547

Mielke, R., Herholz, K., Grond, M., Kessler, J., and Heiss, W. (1994). Clinical deterioration in probable Alzheimer's disease correlates with progressive metabolic impairment of association areas. *Dement. Geriatr. Cogn. Disord.* 5, 36–41. doi: 10.1159/000106692

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)* (Stanford, CA: IEEE), 565–571. doi: 10.1109/3DV.2016.79

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., et al. (2017). Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans. Med. Imaging* 37, 384–395. doi: 10.1109/TMI.2017.2743464

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Singh, V. K., Romani, S., Rashwan, H. A., Akram, F., Pandey, N., Sarker, M. M. K., et al. (2018). "Conditional generative adversarial and convolutional networks for x-ray breast mass segmentation and shape classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Granada: Springer), 833–840. doi: 10.1007/978-3-030-00934-2_92

Smith, G. S., de Leon, M. J., George, A. E., Kluger, A., Volkow, N. D., McRae, T., et al. (1992). Topography of cross-sectional and longitudinal glucose metabolic deficits in Alzheimer's disease: pathophysiologic implications. *Arch. Neurol.* 49, 1142–1150. doi: 10.1001/archneur.1992.00530350056020

Son, J., Park, S. J., and Jung, K.-H. (2017). Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *arXiv preprint arXiv:1706.09318*.

Souly, N., Spampinato, C., and Shah, M. (2017). "Semi supervised semantic segmentation using generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 5688–5696. doi: 10.1109/ICCV.2017.606

Sujji, G. E., Lakshmi, Y., and Jiji, G. W. (2013). MRI brain image segmentation based on thresholding. *Int. J. Adv. Comput. Res.* 3:97.

Szilagyi, L., Benyo, Z., Szilágyi, S. M., and Adam, H. (2003). "MR brain image segmentation using an enhanced fuzzy C-means algorithm," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Cancun: IEEE), 724–726. doi: 10.1109/IEMBS.2003.1279866

Tang, H., Wu, E., Ma, Q., Gallagher, D., Perera, G., and Zhuang, T. (2000). MRI brain image segmentation by multi-resolution edge detection and region selection. *Comput. Med. Imaging Graph.* 24, 349–357. doi: 10.1016/S0895-6111(00)00037-9

Tu, Z., and Bai, X. (2009). Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1744–1757. doi: 10.1109/TPAMI.2009.186

Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., et al. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 396, 1204–1222. doi: 10.1016/S0140-6736(20)30925-9

Yang, H., Lu, X., Wang, S.-H., Lu, Z., Yao, J., Jiang, Y., et al. (2021). Synthesizing multi-contrast MR images via novel 3d conditional variational auto-encoding gan. *Mobile Netw. Appl.* 26, 415–424. doi: 10.1007/s11036-020-01678-1

Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., et al. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage* 108, 214–224. doi: 10.1016/j.neuroimage.2014.12.061

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., and Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med. Image Anal.* 43, 98–111. doi: 10.1016/j.media.2017.10.002

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fnins.2021.729449

F-Value Time-Frequency Analysis: Between-Within Variance Analysis

Hong Gi Yeom[1] *and Hyundoo Jeong*[2]*

[1] Department of Electronics Engineering, Chosun University, Gwangju, South Korea, [2] Department of Mechatronics Engineering, Incheon National University, Incheon, South Korea

Studies on brain mechanisms enable us to treat various brain diseases and develop diverse technologies for daily life. Therefore, an analysis method of neural signals is critical, as it provides the basis for many brain studies. In many cases, researchers want to understand how neural signals change according to different conditions. However, it is challenging to find distinguishing characteristics, and doing so requires complex statistical analysis. In this study, we propose a novel analysis method, FTF (F-value time-frequency) analysis, that applies the F-value of ANOVA to time-frequency analysis. The proposed method shows the statistical differences among conditions in time and frequency. To evaluate the proposed method, electroencephalography (EEG) signals were analyzed using the proposed FTF method. The EEG signals were measured during imagined movement of the left hand, right hand, foot, and tongue. The analysis revealed the important characteristics which were different among different conditions and similar within the same condition. The FTF analysis method will be useful in various fields, as it allows researchers to analyze how frequency characteristics vary according to different conditions.

Keywords: F-value, ANOVA, electroencephalography, analysis of variance, time-frequency analysis

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Román Baravalle,
CONICET Instituto de Física La Plata (IFLP), Argentina
Yufeng Yao,
Changshu Institute of Technology, China

***Correspondence:**
Hyundoo Jeong
hdj@inu.ac.kr

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *23 June 2021*
Accepted: *22 November 2021*
Published: *09 December 2021*

Citation:
Yeom HG and Jeong H (2021) F-Value Time-Frequency Analysis: Between-Within Variance Analysis. Front. Neurosci. 15:729449. doi: 10.3389/fnins.2021.729449

INTRODUCTION

Analysis methods for neural signals are essential tools to understand the mechanisms of brain activity (Rubinov and Sporns, 2010; Yeom et al., 2020a). Neural signal analysis has important academic implications and can be used in various fields, such as medicine, psychology, and biomedical engineering. For instance, the brain area that causes seizure activity can be accurately identified by analyzing neural activity patterns (Jeong et al., 2016; Kim et al., 2021). This is critical to ensure that the patient is free from epilepsy following surgery. Moreover, analysis of neural signals can be used to diagnose conditions such as Alzheimer's disease (Dauwels et al., 2010a,b), attention deficit hyperactivity disorder (ADHD) (Arns et al., 2013; Lenartowicz and Loo, 2014), and schizophrenia (Boutros et al., 2008; Li et al., 2008). Accurate diagnosis of psychological diseases enables prompt and appropriate treatments. Furthermore, movement intention can be predicted by analyzing neural activity (Yeom et al., 2013, 2014, 2020b; Kobler et al., 2020; Mondini et al., 2020; Sosnik and Ben Zur, 2020), and a paralyzed person can control a robot arm according to his or

her intentions (Hochberg et al., 2012; Collinger et al., 2013). Thus, patients who cannot move their bodies can control electric devices according to their thought by analyzing neural signals.

However, analysis of neural signals such as electroencephalography (EEG), magnetic encephalography (MEG), electrocorticography (ECoG), and local field potential (LFP) is quite challenging. Neural activity changes over time and has different characteristics depending on frequency band and brain region (Pfurtscheller and Lopes da Silva, 1999). For example, the power of alpha waves (8 – 13 Hz) and beta waves (13 – 30 Hz) decreases in the right motor cortex during movements of the left hand (Blankertz et al., 2008). After such a movement, the power of the beta waves increases (Yeom et al., 2013). Likewise, the power of alpha and beta waves decreases in the left motor cortex during movements of the right hand (Blankertz et al., 2008). Other movements also cause changes in power in related brain areas (Blankertz et al., 2008). A decrease and increase in power are called event-related desynchronization (ERD) and event-related synchronization (ERS), respectively (Pfurtscheller and Lopes da Silva, 1999). Also, importantly, analysis of neural signals is complicated because, even if the neural signals are measured under the same conditions, they are different each time. Therefore, it is difficult to identify which changes are caused by differences between the conditions or other noise effects.

Time-frequency analysis is a powerful method of analyzing the characteristics of neural activity (Roach and Mathalon, 2008; Tzallas et al., 2009; Fraiwan et al., 2012; Herrmann et al., 2014). Time-frequency analysis visualizes the variation in signal power as a color in both the time and frequency domains. Therefore, the method allows intuitive analysis of the characteristics of neural signals in both the time and frequency domains. The method minimizes the effects of noise by repeatedly measuring neural signals and calculating an average of all trials. However, in this process, transitory or non-phase-locked activities can be ignored (Pfurtscheller and Lopes da Silva, 1999). Time-frequency analysis is a primary and general analysis method widely used in neuroscience studies (Wacker and Witte, 2013; Herrmann et al., 2014). However, using time-frequency analysis, it is difficult to identify which characteristics occur according to different conditions. For example, if time-frequency analysis is performed for various conditions, the analyst needs to find the differences while visually comparing the results.

Although there have been many studies to overcome the disadvantages of the time-frequency analysis (Addison et al., 2009; Cohen, 2019), few studies have applied statistical methods to the time-frequency analysis. Some studies applied the analysis of variance (ANOVA) to the wavelet to evaluate the similarity between the simulation model and the actual system (Atkinson et al., 2017, 2018). However, these studies were aimed to evaluate the simulation, and these methods could not provide information on which characteristics differ among different conditions.

Analyzing differences depending on conditions is critical. The analysis can be used to investigate the cause of diverse diseases or diagnose it (Dubreuil-Vall et al., 2020; Sebastian-Romagosa et al., 2020; Lin et al., 2021). Moreover, it is the core technology in the brain-computer interface (BCI), which controls computers or various electronic devices according to various intentions (Gu and Hua, 2021; Panachakel and Ramakrishnan, 2021). If the characteristics of neural signals change a little under the same condition and a lot among different conditions, the characteristics reflect the conditional changes. This is the basic principle underlying the ANOVA (Scheffé, 1999). In this study, we suggest a new analysis method that can easily identify the neural characteristics that reflect conditional changes by applying the F-value of ANOVA to time-frequency analysis. To the best of our knowledge, the proposed method is the first analysis method that presents the statistical difference among conditions in both time and frequency.

MATERIALS AND METHODS

Data Description

This study used publicly available EEG data from the Laboratory of Brain-Computer Interfaces, Graz University of Technology (Tangermann et al., 2012). The data is available at http://www.bbci.de/competition/iv. Details of the data are described in a previous paper (Tangermann et al., 2012). Briefly, EEG measurements were recorded during four imagined movements (left hand, right hand, foot, and tongue). Nine subjects participated in the study, and the subjects were instructed to imagine movements according to visual stimuli on the screen. EEG was measured using 22-channel Ag/AgCl electrodes. Electrooculography (EOG) was measured using 3-channel unipolar electrodes. The left mastoid was used as the reference and the right mastoid was used as the ground. The sampling rate was 250 Hz. The signals were band-pass filtered at 0.5 to 100 Hz. A notch filter at 50 Hz was applied to remove line noise. To evaluate the F-value time-frequency (FTF) analysis method in another case, we also analyzed steady-state visually evoked potential (SSVEP) data. The analysis results of the SSVEP data are described in the **Supplementary Material**.

Experimental Paradigm

Subjects were seated in a comfortable chair during the experiment. Visual instruction was presented on a computer screen. At the beginning of each session, EOG signals were measured during eyes open, eyes closed, and eye movements. The recording was approximately 5 min. At the beginning of each trial, a fixation cross was displayed at the center of the screen. A short beep sound was presented together. After 2 s, an arrow pointing in one of the four directions appeared for 1.25 s. The fixation cross disappeared 6 s later. Subjects were instructed to imagine movement according to the direction of the arrow until the cross disappeared. The arrow pointing to the left, right, down, and up corresponded to the imagined movement of the left hand, right hand, foot, and tongue, respectively. After the fixation cross, a black screen was presented for a short break. **Figure 1A** illustrates the experimental paradigm. EEG signals were measured during two sessions on different days for each subject as shown in **Figure 1B**. One session is for the training of the prediction

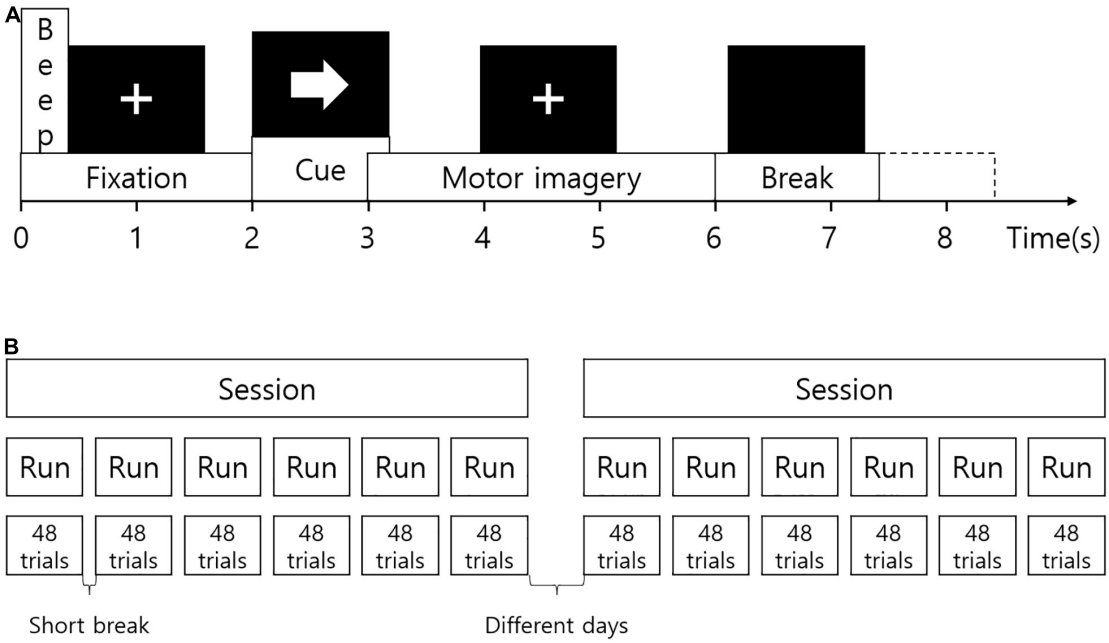


FIGURE 1 | Experimental paradigm. **(A)** The experimental protocol. At the beginning of each trial, a fixation cross was displayed on the screen. A short beep sound was presented together. After 2 s, an arrow pointing in one of the four directions appeared for 1.25 s. The fixation cross disappeared 6 s later. After the fixation cross, a black screen was presented for a short break. **(B)** Configuration of the experiment. EEG signals were measured during two sessions on different days for each subject. One session consisted of 6 runs divided by short breaks. One run included 12 trials for each imagined movement (total, 48 trials).

model, and the other session is for evaluation. The session for training was used for the analysis to find distinguishing features. One session consisted of 6 runs divided by short breaks. One run included 12 trials for each imagined movement (total, 48 trials). Therefore, each movement was imagined in 72 trials in each session.

Time-Frequency Analysis

The proposed analysis method is related to time-frequency analysis. To compare the proposed method with time-frequency analysis method, time-frequency analysis was performed. For the time-frequency analysis, EEG signals were epoched. Epoching is segmentation of EEG signals based on an event. The event corresponds to the time at which visual or audible stimuli are given. In this study, EEG signals were epoched from −2 to 4 s relative to presentation of the arrow. Time-frequency power spectra were calculated for each channel of the epoched EEG data using continuous wavelet transform (CWT). To calculate the CWT, complex Morlet wavelet was used. Time-frequency power spectra were normalized by baseline power for each frequency. Baseline corresponded to the recordings taken from −2 to 0 s relative to the arrow cue. Time-frequency power spectra were averaged by trials. All signal processing was performed using MATLAB, 2020b (Mathworks, Natick, MA, United States).

F-Value Time-Frequency Analysis

We suggest a novel F-value time-frequency analysis method. It visualizes the F-value of ANOVA depending on frequency over time, as shown in **Figure 2**. The F-value is calculated by dividing the variance between groups (among different motor imageries) by the variance within the group (within same imagined movement). A high f-value means a small change within the same condition and a large change among different conditions. In other words, high F-values represent the neural characteristics that vary according to different conditions. F-value time-frequency analysis shows F-values in both the time and frequency domains. Therefore, FTF analysis makes it easy to examine what time and which frequency of the signals are important for classification of the different conditions. The F-value of FTF is calculated as follows:

$$F = \frac{\textit{Between-group variance}}{\textit{Within-group variance}}$$

where between-group variance and within-group variance are calculated as follows:

$$\textit{Between-group variance} = \sum_{i=1}^{K} n_i \frac{(\overline{Y_i} - \overline{Y})^2}{K-1}$$

$$\textit{Within-group variance} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \frac{(Y_{ij} - \overline{Y_i})^2}{N-K}$$

where $[\overline{Y_i}]$ denotes the mean of samples in the i-th class, $[n_i]$ is the number of samples in the i-th group (type of imagined movement),$[\overline{Y}]$ denotes the mean of all 4 groups, $[K]$ denotes the number of groups, $[Y_{ij}]$ is the j-th sample in the i-th group and $[N]$ is the total number of samples. In this study,

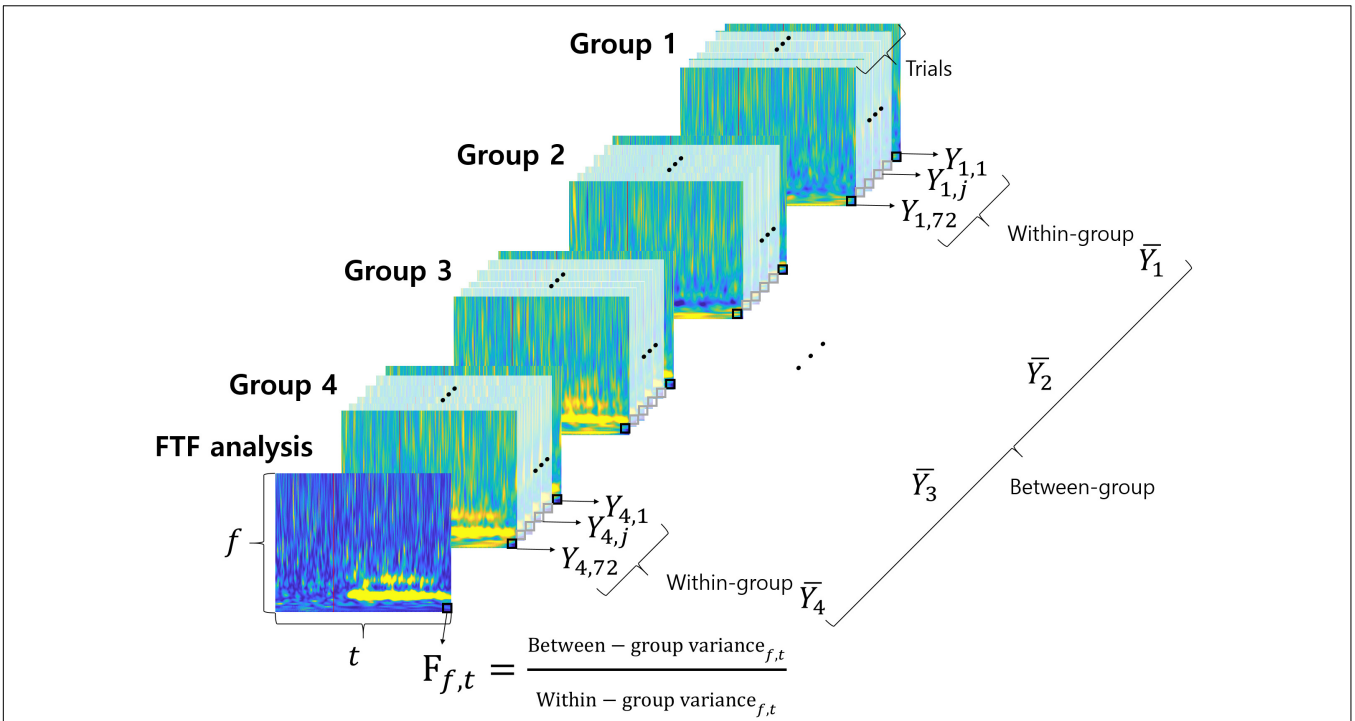


FIGURE 2 | FTF analysis. Each colorful image represents the time-frequency spectrum of a single trial except the front image. The front image shows the results of FTF analysis. Groups 1, 2, 3, and 4 correspond to imagined movements of the left hand, right hand, foot and tongue, respectively. The small black squares are pixels at each time point and frequency.

the parameters were as follows: [n_1=n_2=n_3=n_4= 72], [K= 4], [N= 72∗4 = 288]. The F-values are calculated at each time point and frequency, as follows:

$$F_{f,t}=\frac{Between-group\ variance_{f,t}}{Within-group\ variance_{f,t}}$$

where [f] denotes the frequency and [t] denotes the time point. In this study, f ranged from 1 to 100 and t ranged from 1 to 1,500 (6 s * 250 sampling rate). F-value for the significance level can be determined by the table of the F-distribution (Beyer, 2019). To find the F-value, a significant level should be determined. On the F-distribution table, the F-value for the significance level is the intersection of degrees of freedom (DOF) between-group and DOF of within-group. Instead of the table, an online calculator is available at www.danielsoper.com/statcalc/calculator.aspx?id = 4.

In this study, DOF of between-group and DOF of within-group were 3 (K-1) and 284 (N-K), respectively. The F-value for the significance level is 3.851286 for probability level 0.01 (p = 0.01). **Figure 2** illustrates the FTF analysis method. Each of the colorful figures represents the time-frequency spectrum of a single trial except the front one. The front figure shows the results of FTF analysis. Groups 1, 2, 3, and 4 represent the types of imagined movement (left hand, right hand, foot, and tongue, respectively). The small black squares are the pixels at each time point and frequency. MATLAB code of the FTF analysis is available at https://github.com/honggi82/FTF-analysis.

RESULTS

Figures 3A–D show the averaged time-frequency power spectra for all subjects in channel C3. **Figures 3A–D** correspond to the imagined movement of the left hand, right hand, foot, and tongue, respectively. The red lines show the time points at which arrows were presented on the screen. Blue represents a decrease in power (ERD) and yellow represents an increase in power (ERS) compared to baseline. The time-frequency power spectra is given in arbitrary units (AUs) because the spectra were normalized to the baseline. **Figure 3A** shows ERD at 9 – 40 Hz. **Figure 3B** shows stronger ERD than **Figure 3A** at 9 – 40 Hz. **Figure 3C** reveals a short period of ERD at 15 – 40 Hz and ERS at 10 – 26 Hz after 0.84 s. **Figure 3D** shows the shortest period of ERD at 15 – 22 Hz, ERS at 9 – 17 Hz after 0.54 s, and weak ERS at 18 – 43 Hz until 3 s. **Figures 3A–D** generally reveal ERS at 0 – 8 Hz from 0 to about 3.5 s. It is difficult to determine which frequency causes the differences among different conditions. **Figure 3E** illustrates the FTF analysis in channel C3. The unit of FTF analysis is the F-value. The figure shows significant differences among conditions occur at 8 – 16 Hz and 19 – 42 Hz from 0.55 to 3.36 s. The F-values were low at 0 – 8 Hz, which commonly represented ERS in the time-frequency analysis. The FTF analysis enables examination of which frequency causes the difference. The analysis results mean that ERS of delta and theta waves (0 – 8 Hz) is common among different movements. The analysis also shows that short ERD of beta waves between 0 and 0.5 s is similar among

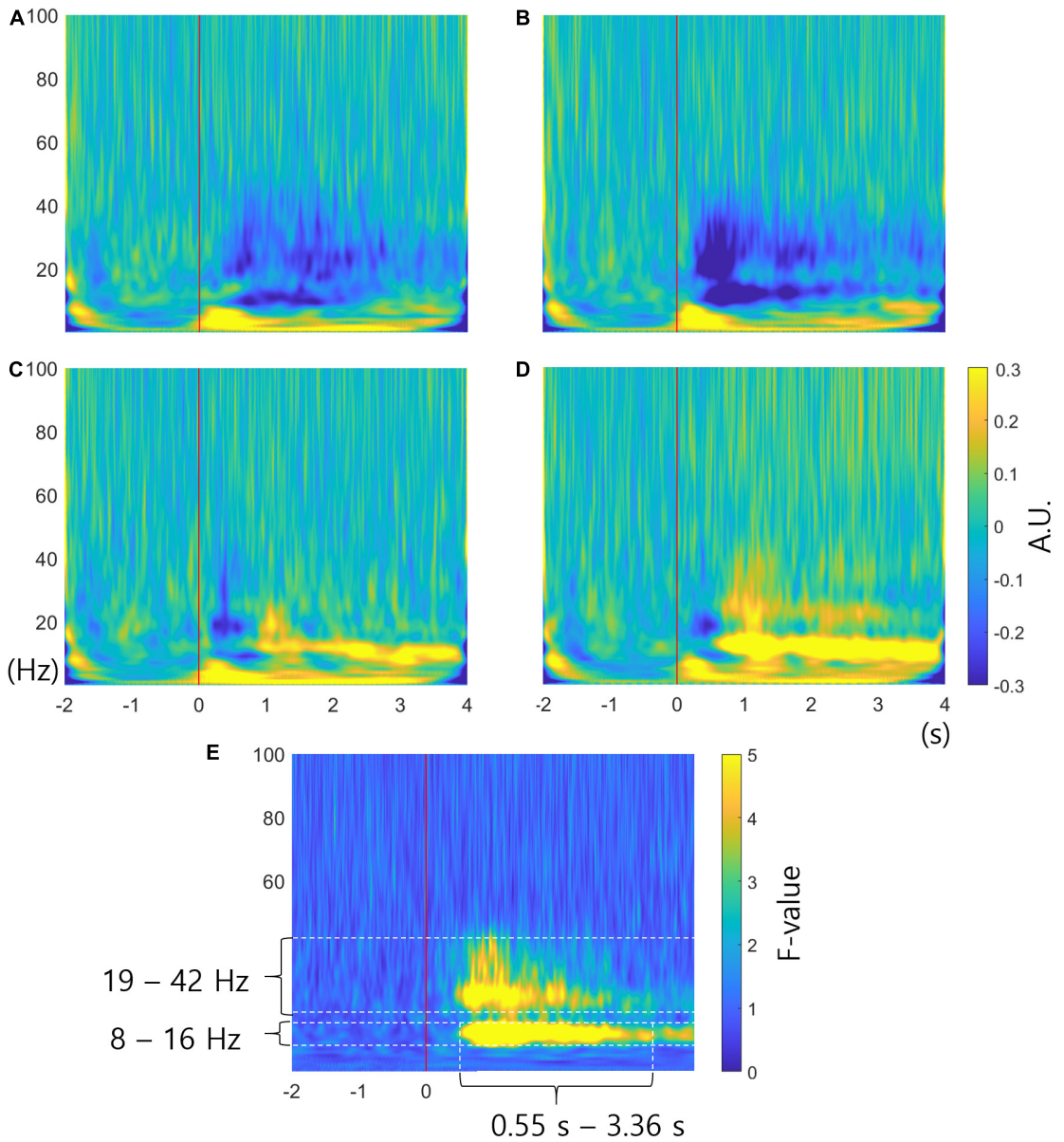


FIGURE 3 | Time-frequency analysis and FTF analysis. **(A–D)** are the averaged time-frequency power spectra for all subjects in channel C3 related to imagined movement of the left hand, right hand, foot and tongue, respectively. The red lines show cue onset. The colors represent the increase or decrease in power in arbitrary units (AUs). **(E)** Averaged FTF analysis for all subjects in channel C3. The unit of FTF analysis is the F-value. The F-value for the significance level is 3.851286 for probability level 0.01 (p = 0.01). It is clear that significant differences among conditions occur at 8 – 16 Hz and 19 – 42 Hz. X-axis, time; y-axis, frequency.

movements. The FTF analysis reveals that ERS or ERD at alpha and beta waves (8 – 42 Hz) are critical features distinguishing different movements.

Figure 4 represents the averaged FTF analysis for all subjects in all channels. The multi-channel FTF analysis enables researchers to examine the overall characteristics at one time easily. Fz, C3, Cz, C4, and Pz are the channel locations in the international 10–20 system. Channels C3, Cz, and C4 show high F-values. Channel Cz is close to the brain area responsible for foot movement. The brain area responsible for right-hand movement is close to channel C3, and the brain area responsible for left-hand movement is close to channel C4. Therefore, FTF analysis reveals the areas that play a distinct role among different movements.

Figure 5A shows the FTF analysis averaged across all subjects and channels. **Figures 5B,C** represent the F-values of the FTF analysis by topography over time at 17 – 27 Hz and 9 – 15 Hz, which have significant F-values in **Figure 5A**. Crucial neural characteristics that differ among the different conditions in frequency, time, or location are readily examined by the FTF analysis. **Figures 5B,C** reveal high F-values in the areas related to the hands and foot.

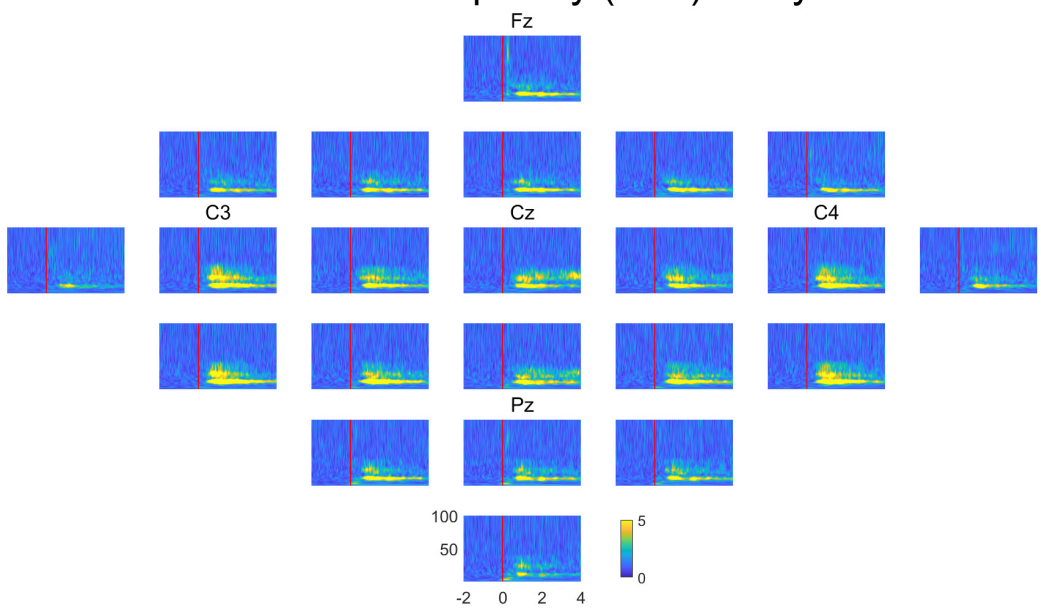


FIGURE 4 | Averaged FTF analysis for all subjects in all channels. Fz, C3, Cz, C4 and Pz are the channel locations in the international 10–20 system. The unit of FTF analysis is the F-value. X-axis, time; y-axis, frequency.

DISCUSSION

F-Value Time-Frequency-Analysis Visualizes Statistical Difference of Neural Signals Among Different Tasks

Although time-frequency analysis is a powerful method of analyzing the characteristics of neural activity, it is difficult to identify which characteristics occur according to the different conditions. On the other hand, the proposed FTF analysis method represents the neural characteristics that vary according to different conditions using statistical values. As shown in **Figures 3A–D**, there are common changes and different changes across conditions. Therefore, it is difficult to examine the difference among tasks intuitively by time-frequency analysis. FTF-analysis showed high F-values when neural activities were different across conditions and low F-values when the neural activities were similar in **Figure 3E**. F-values at 8 – 16 Hz and 19 – 42 Hz were significantly high. The frequencies 8 – 16 Hz and 19 – 42 Hz approximately correspond to alpha and beta waves, respectively. It is well-known that movements of different body parts produce different ERD and ERS patterns at alpha and beta waves (Blankertz et al., 2007, 2008; Tangermann et al., 2012). **Figures 3A–D** represent different ERD and ERS at 8 – 16 Hz and 19 – 42 Hz. **Figures 3A,B** shows weak and strong ERD, respectively. **Figures 3C,D** reveals weak and strong ERS, respectively. Therefore, there are differences across conditions at 8 – 16 Hz and 19 – 42 Hz. On the other hand, F-values at 0 – 8 Hz were low because ERS was common in **Figures 3A–D**. Moreover, ERD was commonly observed at 15 – 20 Hz at 0 – 0.5 s. Therefore, the F-values were low at 15 – 20 Hz at 0 – 0.5 s.

Although F-values were high at broad areas in **Figure 4** because of volume conduction, F-values were higher at channels C3, C4, and Cz than other channels. In the case of EEG, it is difficult to say that the EEG signal at a specific location represents the response of a specific brain area. However, the C3, C4, and Cz channels are close to the brain areas responsible for the right hand, left hand, and foot movements, respectively (Seeck et al., 2017). Moreover, neural responses to the right hand, left hand, and foot movements, are often observed at C3, C4, Cz (Blankertz et al., 2007, 2008; Tangermann et al., 2012). Therefore, it means that FTF analysis shows well the neural characteristics that vary according to different conditions in time, frequency, and channel. The neural characteristics of imagined tongue movements were not clearly observed with time-frequency analysis and FTF analysis. It seems that the tongue-related area of the brain may be more lateral than the measured channels (de Klerk et al., 2015; Wennberg et al., 2019).

Although we analyzed one case of data, the results showed that the proposed FTF analysis effectively represents the differences of neural signals in time, frequency, or area among conditions. We plan to apply FTF analysis to neural signals in various cases through future studies.

Time-Frequency Analysis and F-Value Time-Frequency Analysis Are Complementary

Although FTF analysis identifies neural characteristics that change according to different conditions, it does not mean that the FTF analysis is better than the time-frequency analysis. Time-frequency analysis is especially useful for analyzing neural

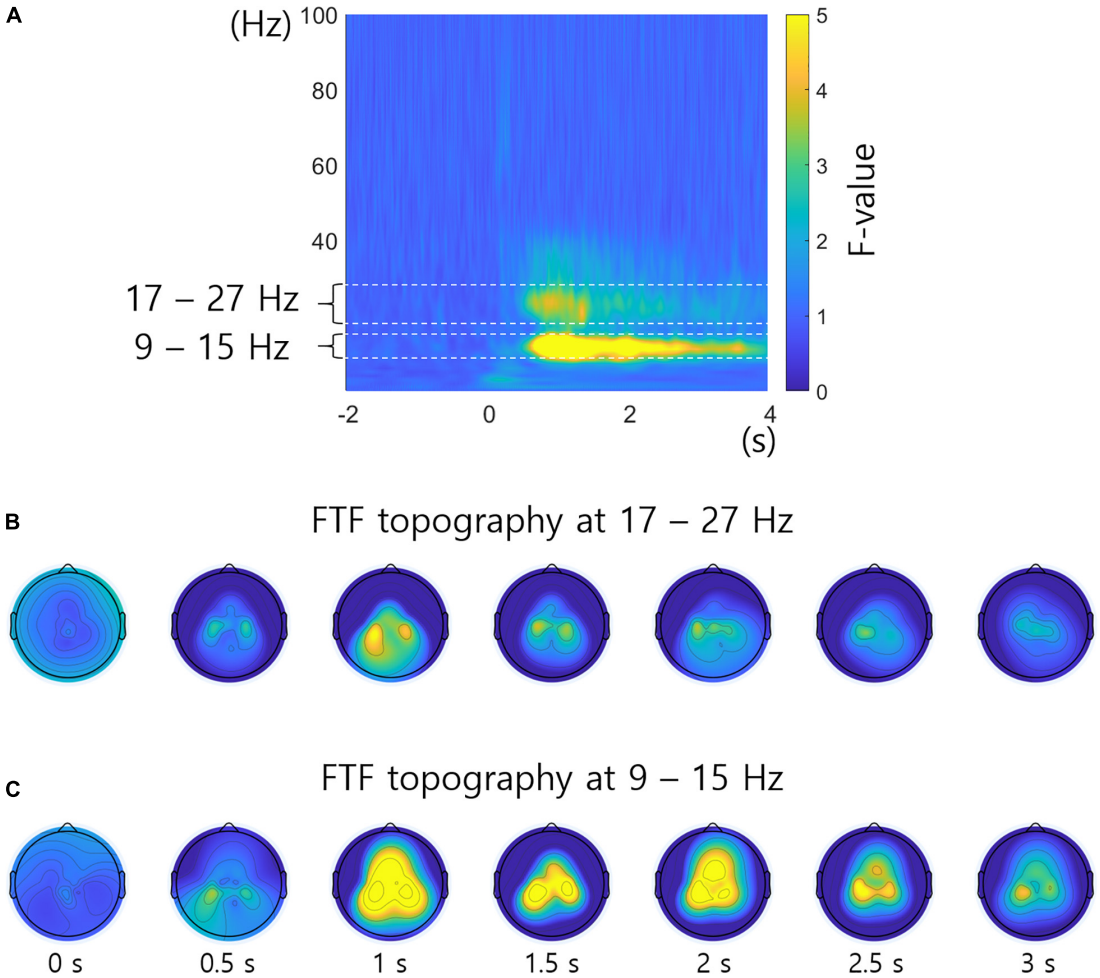


FIGURE 5 | F-value topography in specific frequency bands. **(A)** FTF analysis averaged across all subjects and channels. The 17 – 27 Hz and 9 – 15 Hz frequency bands show significant F-values. **(B)** F-value topography over time at 17 – 27 Hz. **(C)** F-value topography over time at 9 – 15 Hz. The unit of topography is the F-value.

activity under one condition. **Figures 3A–D** show different ERD and ERS patterns depending on the type of imagined movement. Therefore, time-frequency analysis provides valuable information on a specific condition, such as imagined movement of the left hand, whereas FTF is a powerful tool for finding characteristics that differ depending on multiple conditions. Therefore, time-frequency analysis and FTF analysis can be used complementary to each other. FTF analysis can be used in various studies, including those related to the brain but also those related to sound, communication, and so on (Akansu et al., 2010; Varanis et al., 2021).

F-Value Time-Frequency Analysis Has the Disadvantages of Time-Frequency Analysis

F-Value Time-Frequency (FTF) analysis uses time-frequency analysis for F-value calculation. Any time-frequency analysis can be used for the FTF analysis because the F-values are calculated among values of time-frequency analysis. Not only traditional methods but also recently proposed methods also can be applied to the FTF analysis (Cohen, 2019; Yang et al., 2019; Liu et al., 2020; Varanis et al., 2021). However, FTF analysis has the disadvantages of time-frequency analysis because it is based on time-frequency analysis. Time-frequency analysis requires a trade-off between time resolution and frequency resolution (Grochenig, 2001). It means that based on the uncertainty principle, the frequency resolution decreases to increase the time resolution, and the time resolution decreases to increase the frequency resolution. Short-time Fourier transform (STFT) uses fixed resolution in time and frequency (Varanis et al., 2021). However, high-frequency requires high-time resolution and low-frequency requires low-time resolution. CWT increases temporal resolution as frequency increases using a wavelet (Varanis et al., 2021). Time-frequency analysis can be obtained by calculating the absolute values of the CWT. Generally, the CWT's absolute values are calculated for

each trial's signals and then averaged by trials. The process for calculating the absolute values removes the phasor information. Therefore, even if there is important information in the phase, it is difficult to know with time-frequency analysis. Furthermore, the averaging process reduces the transitory or non-phase locked activities, although the process diminishes the noise.

CONCLUSION

In this study, we suggest a novel analysis method that can be used to easily identify neural characteristics that reflect conditional changes by applying the F-value of ANOVA to time-frequency analysis. F-value time-frequency analysis represents the statistical differences among conditions in both the time and frequency domains. EEG signals during 4 movement imagination tasks were analyzed by the FTF method. It was easy to observe critical characteristics that differed in terms of time, frequency, and location. The FTF method will be useful in various fields that analyze how frequency characteristics vary according to different conditions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.bbci.de/competition/iv.

ETHICS STATEMENT

Ethical review and approval were not required for the study on humans 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.

AUTHOR CONTRIBUTIONS

HY proposed and programmed the FTF algorithm. HY and HJ wrote and reviewed the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) and funded by the Ministry of Education (No. 2017R1A6A1A03015496).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.729449/full#supplementary-material

REFERENCES

Addison, P. S., Walker, J., and Guido, R. C. (2009). Time-frequency analysis of biosignals a wavelet transform overview. *IEEE Eng. Med. Biol. Mag.* 28, 14–29.

Akansu, A. N., Serdijn, W. A., and Selesnick, I. W. (2010). Emerging applications of wavelets: a review. *Phys. Commun.* 3, 1–18. doi: 10.1016/j.phycom.2009.07.001

Arns, M., Conners, C. K., and Kraemer, H. C. (2013). A decade of EEG Theta/Beta ratio research in ADHD: a meta-analysis. *J. Attent. Disord.* 17, 374–383. doi: 10.1177/1087054712460087

Atkinson, A. D., Hill, R. R., Pignatiello, J. J., Vining, G. G., White, E. D., and Chicken, E. (2017). Wavelet ANOVA approach to model validation. *Simul. Model. Pract. Theory* 78, 18–27.

Atkinson, A. D., Hill, R. R., Pignatiello, J. J., Vining, G. G., White, E. D., and Chicken, E. (2018). Wavelet ANOVA bisection method for identifying simulation model bias. *Simul. Model. Pract. Theory* 80, 66–74.

Beyer, W. H. (2019). *Handbook of Tables for Probability and Statistics*. CRC Press.

Blankertz, B., Dornhege, G., Krauledat, M., Muller, K. R., and Curio, G. (2007). The non-invasive Berlin brain-computer interface: fast acquisition of effective performance in untrained subjects. *Neuroimage* 37, 539–550. doi: 10.1016/j.neuroimage.2007.01.051

Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., and Muller, K. R. (2008). The Berlin brain-computer interface: accurate performance from first-session in BCI-naive subjects. *IEEE Trans. Biomed. Eng.* 55, 2452–2462. doi: 10.1109/TBME.2008.923152

Boutros, N. N., Arfken, C., Galderisi, S., Warrick, J., Pratt, G., and Iacono, W. (2008). The status of spectral EEG abnormality as a diagnostic test for schizophrenia. *Schizophr. Res.* 99, 225–237. doi: 10.1016/j.schres.2007.11.020

Cohen, M. X. (2019). A better way to define and describe Morlet wavelets for time-frequency analysis. *Neuroimage* 199, 81–86. doi: 10.1016/j.neuroimage.2019.05.048

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., et al. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 381, 557–564. doi: 10.1016/S0140-6736(12)61816-9

Dauwels, J., Vialatte, F., and Cichocki, A. (2010a). Diagnosis of Alzheimer's disease from EEG signals: where are we standing? *Curr. Alzheimer Res.* 7, 487–505. doi: 10.2174/156720510792231720

Dauwels, J., Vialatte, F., Musha, T., and Cichocki, A. (2010b). A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *Neuroimage* 49, 668–693. doi: 10.1016/j.neuroimage.2009.06.056

de Klerk, C. C. J. M., Johnson, M. H., and Southgate, V. (2015). An EEG study on the somatotopic organisation of sensorimotor cortex activation during action execution and observation in infancy. *Dev. Cogn. Neurosci.* 15, 1–10. doi: 10.1016/j.dcn.2015.08.004

Dubreuil-Vall, L., Ruffini, G., and Camprodon, J. A. (2020). Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. *Front. Neurosci.* 14:251. doi: 10.3389/fnins.2020.00251

Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H., and Dickhaus, H. (2012). Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Comput. Methods Prog. Biomed.* 108, 10–19. doi: 10.1016/j.cmpb.2011.11.005

Grochenig, K. (2001). *Foundations of Time-Frequency Analysis: With 15 Figures*. Boston: Birkhauser.

Gu, Y., and Hua, L. (2021). A novel smart motor imagery intention human-computer interaction model using extreme learning machine and EEG signals. *Front. Neurosci.* 15:685119. doi: 10.3389/fnins.2021.685119

Herrmann, C. S., Rach, S., Vosskuhl, J., and Struber, D. (2014). Time-frequency analysis of event-related potentials: a brief tutorial. *Brain Topogr.* 27, 438–450. doi: 10.1007/s10548-013-0327-5

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372–375. doi: 10.1038/nature11076

Jeong, W., Kim, J. S., and Chung, C. K. (2016). Usefulness of multiple frequency band source localizations in ictal MEG. *Clin. Neurophysiol.* 127, 1049–1056. doi: 10.1016/j.clinph.2015.07.015

Kim, D., Kim, J. S., Jeong, W., Shin, M.-S., and Chung, C. K. (2021). Critical area for memory decline after mesial temporal resection in epilepsy patients. *J. Neurosurg.* 134, 659–677.

Kobler, R. J., Almeida, I., Sburlea, A. I., and Muller-Putz, G. R. (2020). Using machine learning to reveal the population vector from EEG signals. *J. Neural Eng.* 17:026002. doi: 10.1088/1741-2552/ab7490

Lenartowicz, A., and Loo, S. K. (2014). Use of EEG to diagnose ADHD. *Curr. Psychiatry Rep.* 16:498.

Li, Y. J., Tong, S. B., Liu, D., Gai, Y., Wang, X. Y., Wang, J. J., et al. (2008). Abnormal EEG complexity in patients with schizophrenia and depression. *Clin. Neurophysiol.* 119, 1232–1241. doi: 10.1016/j.clinph.2008.01.104

Lin, N., Gao, J., Mao, C. H., Sun, H. Y., Lu, Q., and Cui, L. Y. (2021). Differences in multimodal electroencephalogram and clinical correlations between early-onset Alzheimer's disease and frontotemporal dementia. *Front. Neurosci.* 15:687053. doi: 10.3389/fnins.2021.687053

Liu, F. S., Gao, S. J., Tian, Z., and Liu, D. Z. (2020). A new time-frequency analysis method based on single mode function decomposition for offshore wind turbines. *Mar. Struct.* 72:102782. doi: 10.1016/j.marstruc.2020.102782

Mondini, V., Kobler, R. J., Sburlea, A. I., and Muller-Putz, G. R. (2020). Continuous low-frequency EEG decoding of arm movement for closed-loop, natural control of a robotic arm. *J. Neural Eng.* 17:046031. doi: 10.1088/1741-2552/aba6f7

Panachakel, J. T., and Ramakrishnan, A. G. (2021). Decoding covert speech from EEG-a comprehensive review. *Front. Neurosci.* 15:642251. doi: 10.3389/fnins.2021.642251

Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Roach, B. J., and Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophr. Bull.* 34, 907–926. doi: 10.1093/schbul/sbn093

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003

Scheffé, H. (1999). *The Analysis of Variance*. New York, NY: Wiley-Interscience Publication.

Sebastian-Romagosa, M., Udina, E., Ortner, R., Dinares-Ferran, J., Cho, W., Murovec, N., et al. (2020). EEG biomarkers related with the functional state of stroke patients. *Front. Neurosci.* 14:582. doi: 10.3389/fnins.2020.00582

Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., et al. (2017). The standardized EEG electrode array of the IFCN. *Clin. Neurophysiol.* 128, 2070–2077. doi: 10.1016/j.clinph.2017.06.254

Sosnik, R., and Ben Zur, O. (2020). Reconstruction of hand, elbow and shoulder actual and imagined trajectories in 3D space using EEG slow cortical potentials. *J. Neural Eng.* 17:016065. doi: 10.1088/1741-2552/ab59a7

Tangermann, M., Muller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6:55. doi: 10.3389/fnins.2012.00055

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans. Inform. Technol. Biomed.* 13, 703–710. doi: 10.1109/TITB.2009.2017939

Varanis, M., Silva, A. L., Balthazar, J. M., and Pederiva, R. (2021). A tutorial review on time-frequency analysis of non-stationary vibration signals with nonlinear dynamics applications. *Braz. J. Phys.* 51, 859–877.

Wacker, M., and Witte, H. (2013). Time-frequency techniques in biomedical signal analysis a tutorial review of similarities and differences. *Methods Inform. Med.* 52, 279–296.

Wennberg, R., Dominguez, L. G., and Del Campo, J. M. (2019). Magnetoencephalographic source localization of the eye area of the motor homunculus. *Can. J. Neurol. Sci.* 46, 108–114. doi: 10.1017/cjn.2018.373

Yang, Y., Peng, Z. K., Zhang, W. M., and Meng, G. (2019). Parameterised time-frequency analysis methods and their engineering applications: a review of recent advances. *Mech. Syst. Signal Process.* 119, 182–221.

Yeom, H. G., Kim, J. S., and Chung, C. K. (2013). Estimation of the velocity and trajectory of three-dimensional reaching movements from non-invasive magnetoencephalography signals. *J. Neural Eng.* 10:026006. doi: 10.1088/1741-2560/10/2/026006

Yeom, H. G., Kim, J. S., and Chung, C. K. (2014). High-accuracy brain-machine interfaces using feedback information. *PLoS One* 9:e0103539. doi: 10.1371/journal.pone.0103539

Yeom, H. G., Kim, J. S., and Chung, C. K. (2020a). Brain mechanisms in motor control during reaching movements: transition of functional connectivity according to movement states. *Sci. Rep.* 10:567. doi: 10.1038/s41598-020-57489-7

Yeom, H. G., Kim, J. S., and Chung, C. K. (2020b). LSTM improves accuracy of reaching trajectory prediction from magnetoencephalography signals. *IEEE Access* 8, 20146–20150.

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 16 December 2021
doi: 10.3389/fnins.2021.801008

3D U-Net Improves Automatic Brain Extraction for Isotropic Rat Brain Magnetic Resonance Imaging Data

Li-Ming Hsu[1,2,3†], Shuai Wang[4†], Lindsay Walton[1,2,3], Tzu-Wen Winnie Wang[1,2,3], Sung-Ho Lee[1,2,3] and Yen-Yu Ian Shih[1,2,3*]*

[1] Center for Animal Magnetic Resonance Imaging, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2] Biomedical Research Imaging Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [3] Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [4] School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China

Brain extraction is a critical pre-processing step in brain magnetic resonance imaging (MRI) analytical pipelines. In rodents, this is often achieved by manually editing brain masks slice-by-slice, a time-consuming task where workloads increase with higher spatial resolution datasets. We recently demonstrated successful automatic brain extraction via a deep-learning-based framework, U-Net, using 2D convolutions. However, such an approach cannot make use of the rich 3D spatial-context information from volumetric MRI data. In this study, we advanced our previously proposed U-Net architecture by replacing all 2D operations with their 3D counterparts and created a 3D U-Net framework. We trained and validated our model using a recently released CAMRI rat brain database acquired at isotropic spatial resolution, including T2-weighted turbo-spin-echo structural MRI and T2*-weighted echo-planar-imaging functional MRI. The performance of our 3D U-Net model was compared with existing rodent brain extraction tools, including Rapid Automatic Tissue Segmentation, Pulse-Coupled Neural Network, SHape descriptor selected External Regions after Morphologically filtering, and our previously proposed 2D U-Net model. 3D U-Net demonstrated superior performance in Dice, Jaccard, center-of-mass distance, Hausdorff distance, and sensitivity. Additionally, we demonstrated the reliability of 3D U-Net under various noise levels, evaluated the optimal training sample sizes, and disseminated all source codes publicly, with a hope that this approach will benefit rodent MRI research community.

Significant Methodological Contribution: We proposed a deep-learning-based framework to automatically identify the rodent brain boundaries in MRI. With a fully 3D convolutional network model, 3D U-Net, our proposed method demonstrated improved performance compared to current automatic brain extraction methods, as shown in several qualitative metrics (Dice, Jaccard, PPV, SEN, and Hausdorff). We trust that this tool will avoid human bias and streamline pre-processing steps during 3D high resolution rodent brain MRI data analysis. The software developed herein has been disseminated freely to the community.

Keywords: 3D U-Net, MRI, brain extraction, segmentation, brain mask, rat brain

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jan Egger,
Graz University of Technology, Austria
Yang Lei,
Emory University, United States

****Correspondence:***
Li-Ming Hsu
limingh@email.unc.edu
Yen-Yu Ian Shih
shihy@unc.edu

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *24 October 2021*
Accepted: *15 November 2021*
Published: *16 December 2021*

Citation:
Hsu L-M, Wang S, Walton L, Wang T-WW, Lee S-H and Shih Y-YI (2021) 3D U-Net Improves Automatic Brain Extraction for Isotropic Rat Brain Magnetic Resonance Imaging Data. Front. Neurosci. 15:801008. doi: 10.3389/fnins.2021.801008

INTRODUCTION

Magnetic resonance imaging (MRI) is a commonly utilized technique to noninvasively study the anatomy and function of rodent brains (Mandino et al., 2019). Among the data pre-processing procedures, brain extraction is an important step that ensures the success of subsequent registration processes (Uhlich et al., 2018). Automating brain extraction (a.k.a. skull stripping) is particularly challenging for rodent brains as compared to humans because of differences in brain/scalp tissue geometry, image resolution with respect to brain-scalp distance, tissue contrast around the skull, and sometimes signal artifacts from surgical manipulations (Hsu et al., 2020). Additionally, rodent brain MRI data is typically acquired at higher magnetic fields (mostly > 7T), where stronger susceptibility artifacts and field biases represent challenges to the rodent brain extraction process. Further, most rodent MRI studies do not utilize a volume receiver and therefore exhibit higher radiofrequency (RF) coil inhomogeneity. For these reasons, the extraction tools that work with human brains (Kleesiek et al., 2016; Fatima et al., 2020; Wang et al., 2021) cannot be directly adopted for rodent brain applications. In practice, rodent brain extraction is often achieved by manually drawing brain masks for every slice, making it a time-consuming and operator-dependent process. Datasets with high through-plane resolution make manual brain extraction a particularly daunting task. Therefore, a robust and reliable automatic brain extraction tool would streamline the pre-processing pipeline, avoid personnel bias, and significantly improve research efficiency (Babalola et al., 2009; Lu et al., 2010; Gaser et al., 2012; Feo and Giove, 2019; Hsu et al., 2020).

To date, the most prominent tools to address rodent MRI brain extraction include Pulse-Coupled Neural Network (PCNN)-based brain extraction proposed by Chou et al. (2011), Rapid Automatic Tissue Segmentation (RATS) proposed by Oguz et al. (2014), and SHape descriptor selected External Regions after Morphologically filtering (SHERM) proposed by Liu et al. (2020), as well as a convolutional deep-learning based algorithm, 2D U-Net, proposed by Hsu et al. (2020). While PCNN, RATS, and SHERM have demonstrated remarkable success [detailed introduction and comparisons discussed in Hsu et al. (2020)], their performance is subject to brain size, shape, texture, and contrast; hence, their settings often need to be adjusted per MRI-protocol for optimal results. In contrast, the U-Net algorithm explores and learns the hierarchical features from the training dataset, and provides a user-friendly and more universally applicable platform (Ronneberger et al., 2015; Yogananda et al., 2019).

As a specific type of convolutional neural network (CNN) architecture (Krizhevsky et al., 2012), U-Net has proven valuable in biomedical image segmentation (Ronneberger et al., 2015; Yogananda et al., 2019). U-Net utilizes the encoder/decoder structure that easily integrates multi-scale information and has better gradient propagation during training. Our previous 2D U-Net approach uses 2D convolutional kernels to predict brain boundaries on a single slice basis. However, since the 2D framework only takes a single slice as input and does not utilize information across slice direction, it inherently fails to leverage context from adjacent slices in a volumetric MRI dataset. To improve upon and possibly outperform 2D U-Net, a 3D U-Net framework using 3D convolutional kernels to predict segmentation predictions on volumetric patches must be explored (Çiçek et al., 2016).

In this work, we demonstrated the use of 3D U-Net for brain extraction in high resolution 3D rat brain MRI data. The whole network is implemented based on Keras (Chollet, 2015) with TensorFlow (Martín et al., 2016) as the backend. We trained and tested the 3D U-Net model for brain extraction performance using a recently released CAMRI rat brain database (Lee et al., 2021), including T2-weighted (T2w) rapid acquisition with relaxation enhancement (RARE) structural MRI (0.2 mm isotropic resolution) and T2*-weighted (T2*w) echo-planar-imaging (EPI) functional MRI (0.4 mm isotropic resolution). The 3D U-Net model was compared with existing rodent brain extraction tools, including RATS, PCNN, SHERM, and 2D U-Net. To benchmark the utility of this approach to rodent MRI data of various quality, we assessed 3D U-Net performance under various noise levels and evaluated the optimal training sample sizes. It is our hope that the information provided herein, together with dissemination of source codes, will benefit the rodent MRI research community.

MATERIALS AND METHODS

Dataset Descriptions

This study utilizes a recently disseminated CAMRI dataset (Lee et al., 2021), available at https://doi.org/10.18112/openneuro.ds003646.v1.0.0. The CAMRI dataset consisted of rats aged P80–210 weighing 300–450 g across cohorts of three different rat strains: Sprague-Dawley (n = 41), Long-Evans (n = 13), and Wistar (n = 33); and both sexes: male (n = 65) and female (n = 22). Detailed information about each cohort can be found in Lee et al. (2021). Each subject contains a T2w RARE and an T2*w EPI, and resolutions were 0.2 mm isotropic and 0.4 mm isotropic, respectively. To train our U-Net model, we first established a training dataset by randomly selecting 80% of the T2w RARE and T2*w EPI images in the CAMRI rat database (69 subjects), leaving the remaining 20% of the data as final performance testing dataset (18 subjects). In the training process, we further randomly selected 80% of the data from the training dataset (55 subjects). The remaining 20% of the data from the training dataset was used to validate the training of U-Net model (14 subjects). We repeated the training-validation process five times to avoid potential bias in data splitting. For each U-Net algorithm, the U-Net model with the highest averaged validation accuracy was then used as the final model for testing (**Supplementary Figure 2**).

U-Net

The 3D U-Net framework is shown in **Figure 1**. The contracting path includes 32 feature maps in the first convolutional block, 64 in the second, then 96, 128, and 256 in the third, fourth,

and fifth, respectively. Compared to the configuration described by Ronneberger et al. (2015), we replaced the cross-entropy loss function with the Dice coefficient loss (Wang et al., 2020) to free the optimization process from a class-imbalance problem (Milletari et al., 2016). Since we evaluated T2w RARE and T2*w EPI data, we performed spatial normalization for distinct resolutions. For spatial normalization, we resampled all images into the same spatial resolution at 0.2 mm × 0.2 mm × 0.2 mm using nearest-neighbor (NN) interpolation. The NN was chosen because it is most suitable for binary images (brain mask). In U-Net training, the voxels belonging to the rat brain were labeled as 1 and other voxels (background) were labeled as 0. Our network was implemented using Keras (Chollet, 2015) with TensorFlow (Martín et al., 2016) as the backend. The initial learning rate and batch size were $1e^{-3}$ and 16, respectively. We used Adam (Kingma and Ba, 2015) as the optimizer and clipped all parameter gradients to a maximum norm of 1. During the training process, we randomly cropped 64 × 64 × 64 sized patches from all directions as the input. Patches were chosen to be 64 × 64 × 64 to maintain maximum patch size and cover as much of the rat brain as possible (∼2,000 mm^3) without requiring excessive computational demand. During the inference process, the extracted and overlapped patches were input into the trained model with a 16 × 16 × 16 stride. The overlapped predictions were averaged and then resampled back to the original resolution using nearest-neighbor interpolation to generate final output. We also trained (1) 2D U-Net with patch size of 64 × 64 (2D U-Net64) and a 16 × 16 stride (Hsu et al., 2020) and (2) 3D U-Net with patch size of 16 × 16 × 16 and a 4 × 4 × 4 stride (3D U-Net16) to compare the segmentation performance with the proposed 3D U-Net of patch size 64 × 64 × 64 and a 16 × 16 × 16 stride (3D U-Net64).

Reproducibility of 3D U-Net Performance

To evaluate the 3D U-Net64 performance reproducibility, we reexamined its accuracy under two conditions: (1) Adding different Gaussian noise in the testing images, and (2) using a different number of rats from the original cohort in the training process. Specifically, as for the first validation analysis, we added Gaussian white noise in the normalized testing images with variance from 5×10^{-4} to 5×10^{-5} with step 5×10^{-5} to investigate the segmentation performance. For each testing image, the signal-to-noise ratio (SNR) was estimated to represent the image noise levels by calculating the ratio of signal intensities in the area of interest and the background (**Supplementary Figure 1**). We estimated the average signal by using two spherical volumes of interest (VOI) with diameter = 1 mm in bilateral striatum. We also put two VOIs at background to extract noise in standard deviation of signal. To evaluate the optimal subject number in 3D U-Net64 training process, we randomly selected 5–55 training subjects in increments of 5 subjects from the total 55 training datasets, and 2, 8, and 14 validation subjects from the total 14 validation datasets. This random selection was repeated 5 times to avoid bias. We then calculated accuracy of the derived brain masks across each testing dataset.

Evaluation Methods

To demonstrate the reliability of our proposed method, we compared our 3D U-Net method with the most prominently used methods for rat brain segmentation: RATS (Oguz et al., 2014), PCNN (Chou et al., 2011), SHERM (Liu et al., 2020), and 2D U-Net (Hsu et al., 2020). All images were bias-corrected for field inhomogeneities using Advanced Normalization Tools (N4ITK) (Avants et al., 2009; Tustison et al., 2010). The parameters in each method were chosen according to the best parameters suggested in the literature. For the RATS algorithm, the intensity threshold (T) was set to the average intensity in the entire image and the brain size value (Vt) was set to 1,650 mm^3 (Oguz et al., 2014). For the PCNN algorithm, the brain size range was set to 1,000–3,000 mm^3 (Chou et al., 2011). For SHERM, the brain size range was set to 500–1,900 mm^3 (Liu et al., 2020). The convexity threshold in SHERM, defined as the ratio between the volume of a region and that of its convex hull, was set to 0.5.

To quantitatively evaluate the segmentation performance of 3D U-Net64, 2D U-Net64, 3D U-Net16, RATS, PCNN, and SHERM, we estimated the similarity of the generated brain segmentation results compared to manual brain masks (ground truth) drawn by an anatomical expert according to the Paxinos and Watson (2007) rat atlas. The manual segmentation was performed at the original MRI resolution before data resampling to 0.2 mm × 0.2 mm × 0.2 mm for U-Net training. Evaluations included: (1) volumetric overlap assessments via Dice, the similarity of two samples; (2) Jaccard, the similarity of two samples where Dice doesn't satisfy the triangle inequality; (3) positive predictive value (PPV), the rate of true positives in prediction results; (4) sensitivity (SEN), the rate of true positives in manual delineation; (5) center-of-mass distance (CMD), the Euclidean distance to the center of mass of two samples; and (6) a surface distance assessment by Hausdorff distance, the distance of two samples. The following definitions were used for each: $Dice = 2(|A \cap B|) \,/\, (|A| + |B|)$, $Jaccard = (|A \cap B|) \,/\, (|A \cup B|)$, $PPV = (|A \cap B|) \,/ B$, $SEN = (|A \cap B|) \,/ A$, and $Hausdorff = max\left\{h\left(A, B\right), h(B, A)\right\}$ and $h(A, B) = \underset{a \in A}{max}\left\{\underset{b \in B}{min}\, d(a, b)\right\}$ where A denotes the voxel set of the manually delineated volume, B denotes the voxel set of the predicted volume, and $d(a, b)$ is the Euclidian distance between a and b. The Hausdorff distance was only estimated in-plane to avoid confounds from non-uniformly sampled data. The maximal Hausdorff distance (i.e., worst matching) across slices for each subject was then used for comparison. Superior performance was indicated by higher Dice, Jaccard, PPV, and SEN, and lower Hausdorff values. We also reported the computation time on a Linux-based [Red Hat Enterprise Linux Server release 7.4 (Maipo)] computing system (Intel E5-2680 v3 processor, 2.50 GHz, 256-GB RAM) for each method. The computation times reported do not include pre-processing steps (i.e., signal normalization, image resampling, and bias correction). Paired t-tests were used for statistical comparisons between different algorithms, and two-sample t-tests were used to compare T2w RARE and T2*w EPI images in each algorithm. The threshold for significance was set to the alpha level ($p < 0.05$).

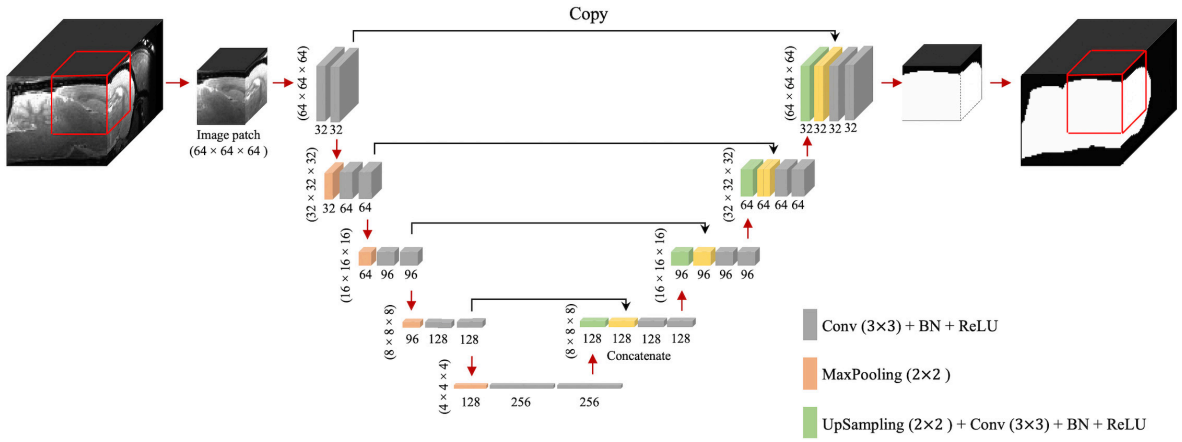


FIGURE 1 | 3D U-Net architecture. Boxes represent cross-sections of square feature maps. Individual map dimensions are indicated on lower left, and number of channels are indicated below the dimensions. The leftmost map is a 64 × 64 × 64 normalized MRI image patched from the original MRI map, and the rightmost represents binary ring mask prediction. Red arrows represent operations specified by the colored box, while black arrows represent copying skip connections. Conv, convolution; BN, batch normalization; ReLU, rectified linear unit.

RESULTS

Figure 2 illustrates the performance of our trained 3D U-Net64 algorithm compared to 2D U-Net64, 3D U-Net16, RATS, PCNN, and SHERM for rat brain segmentation in the CAMRI dataset. Across all measures except PPV (which had no significant differences), 3D U-Net64 showed superior T2w RARE brain segmentation performance over other existing methods (RATS, PCNN, and SHERM). Notably, although all the U-Net-based approaches produced ideal results with Dice > 0.90, the 3D U-Net64 showed significantly higher accuracy ($p < 0.05$) versus the remaining U-Net approaches (2D U-Net64 and 3D U-Net16). The high PPV (>0.90) and low SEN (<0.90) from other existing methods (RATS, PCNN, and SHERM) indicate that brain segmentation was underestimated. In contrast, the low PPV (0.91 on anisotropic T2w RARE and 0.87 on anisotropic T2*w EPI) and high SEN (>0.95) from 2D U-Net16 suggest an overestimation. The significantly lower Hausdorff distance (4.29 on T2w RARE and 4.71 on T2*w EPI) and CMD (0.24 on T2w RARE and 0.25 on T2*w EPI) in 3D U-Net64 further supports its superior ability to match the ground-truth. Notably, all U-Net methods provide excellent (Dice > 0.95) or high (Dice > 0.90) accuracy on T2w RARE and T2*w EPI images. All methods showed significant lower Dice in T2*w EPI images as compared to T2w RARE images ($p < 0.05$). Specifically, the 3D U-Net64 still reached outstanding T2*w EPI brain segmentation performance metrics (Dice, Jaccard, PPV, and SEN > 0.95), whereas RATS, PCNN, and SHERM had lower T2*w EPI brain segmentation metrics (Dice, Jaccard, and SEN < 0.90). The compromised performance in the T2*w EPI image compared with T2w RARE indicates that these methods are less effective at handling data with lower spatial resolution. Further, 3D U-Net64 showed high accuracy (Dice > 0.95) in a validation study (**Supplementary Figure 2**). Together, these results suggest that 3D U-Net64 is a reliable and reproducible approach for rat brain segmentation.

Figures 3, **4** illustrate the best and worst Dice score cases on T2w RARE images from the CAMRI dataset using all six methods. In the best case, both 2D U-Net64 and 3D U-Net16 provided nearly perfect segmentation (Dice > 0.98), and although the RATS, PCNN, and SHERM algorithms also showed high performance (Dice = 0.95), the inferior brain boundaries were less accurate. In the worst case, all U-Net methods still achieved a satisfactory segmentation with Dice > 0.90, but the RATS, PCNN, and SHERM algorithms failed to identify the brainstem, olfactory bulb, and inferior brain regions where the MRI signal was weaker.

Figure 5 illustrates the best and worst Dice score cases on T2*w EPI from the CAMRI dataset using all six methods. In the best case, all U-Net methods provide nearly perfect segmentation (Dice > 0.98). The RATS (Dice = 0.89), PCNN (Dice = 0.93), and SHERM (Dice = 0.83) algorithms showed acceptable segmentation performance but underestimated inferior brain regions. In the worst case, all U-Net methods still achieve a satisfactory segmentation with Dice > 0.90, but the RATS, PCNN, and SHERM algorithms failed to accurately identify the inferior brain boundaries and incorporated excessive tissue outside the brain (Dice < 0.85).

We performed two validation analyses to evaluate the reliability of 3D U-Net segmentation performance. First, we added Gaussian noise to each testing image to evaluate 3D U-Net64 brain segmentation under different noisy environments. **Figure 6** shows that the SNR and 3D U-Net64 segmentation performance decrease with increasing Gaussian white noise variance. The original mean SNR is about 26 for both T2w RARE and T2*w EPI data. When the SNR > 10 (i.e., reduced to 38% of original SNR) in T2w RARE images and SNR > 14 (i.e., reduced to 54% of original SNR) in T2*w EPI images, the segmentation accuracy could still reach an excellent Dice > 0.95. **Supplementary Figures 3**, **4** illustrate the best and worst examples on original T2w RARE and T2*w EPI images.

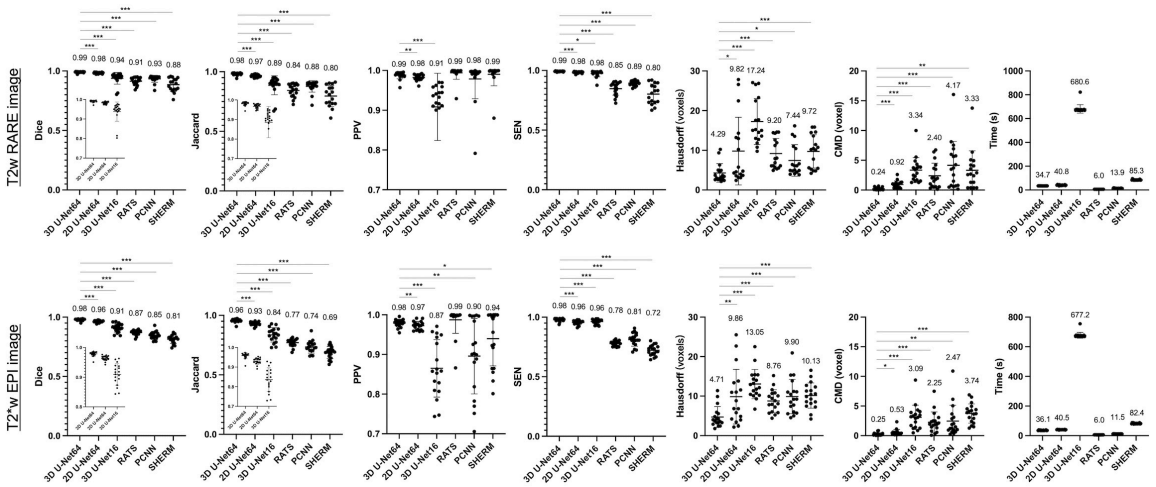


FIGURE 2 | Brain segmentation performance metrics for 3D U-Net64, 2D U-Net64, 3D U-Net16, RATS, PCNN, and SHERM on the CAMRI T2w RARE (*upper row*) and T2*w EPI (*lower row*) data. Average value is shown above each bar. Two-tailed paired *t*-tests were used for statistical comparison between 3D U-Net64 with other methods (*$p < 0.05$, **$p < 0.01$, and *** $p < 0.001$).

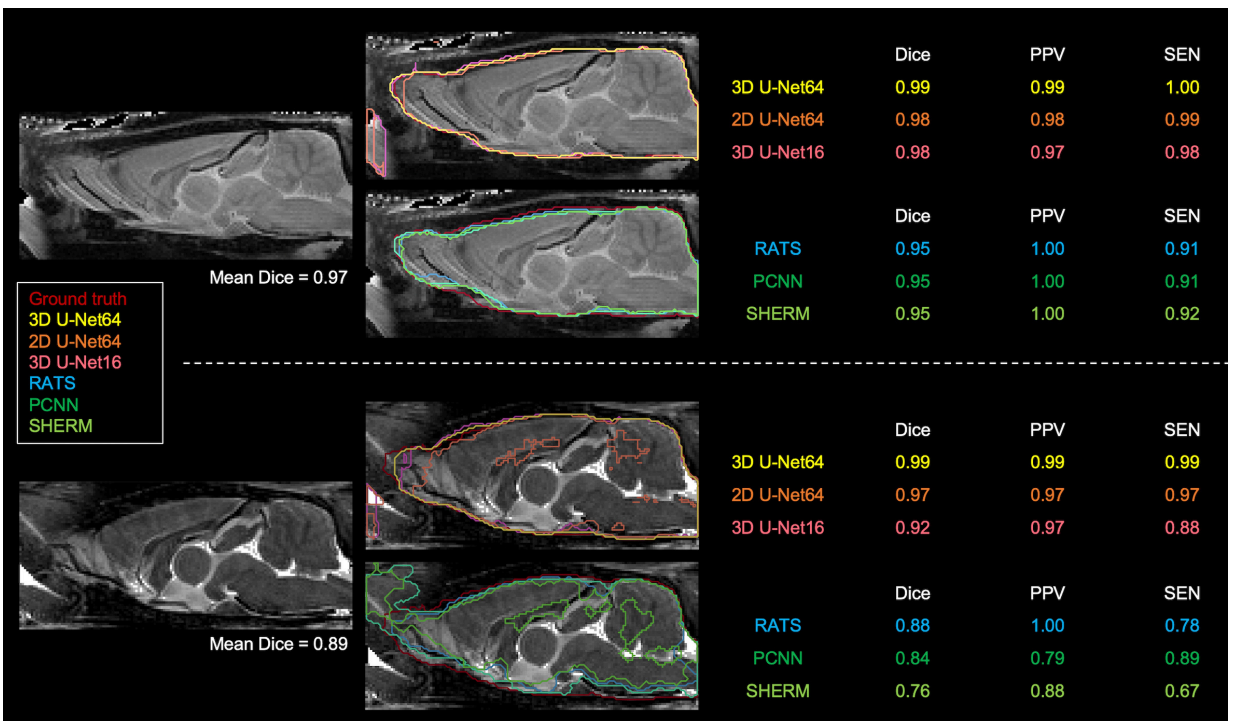


FIGURE 3 | Best (*upper panel*) and worst (*lower panel*) segmentation comparisons for T2w RARE images. Selection was based on the highest and lowest mean Dice score (listed below the brain map) averaged over the six methods (3D U-Net64, 2D U-Net64, 3D U-Net16, RATS, PCNN, and SHERM). Anterior and inferior slices are more susceptible to error in RATS, PCNN, and SHERM, whereas all U-Net algorithms yield high similarly to the ground truth (all Dice > 0.90).

Next, we evaluated the performance of 3D U-Net64 with regards to the model-training and model-validating sample sizes used during the model training process. We re-trained the 3D U-Net64 model with randomly selected training subgroups. Compared to the "standard model" trained on 55 subjects and validated on 14 subjects, the model with training rats < 25 showed significantly decreased accuracy for T2w RARE images, and the model with training rats < 50 showed significantly decreased accuracy for T2*w EPI images (**Figure 7**). However, it is worth noting that the model still reached excellent and stable segmentation performance (Dice > 0.95) with training subjects ≥ 10 for both T2w RARE and T2*w EPI MRI data.

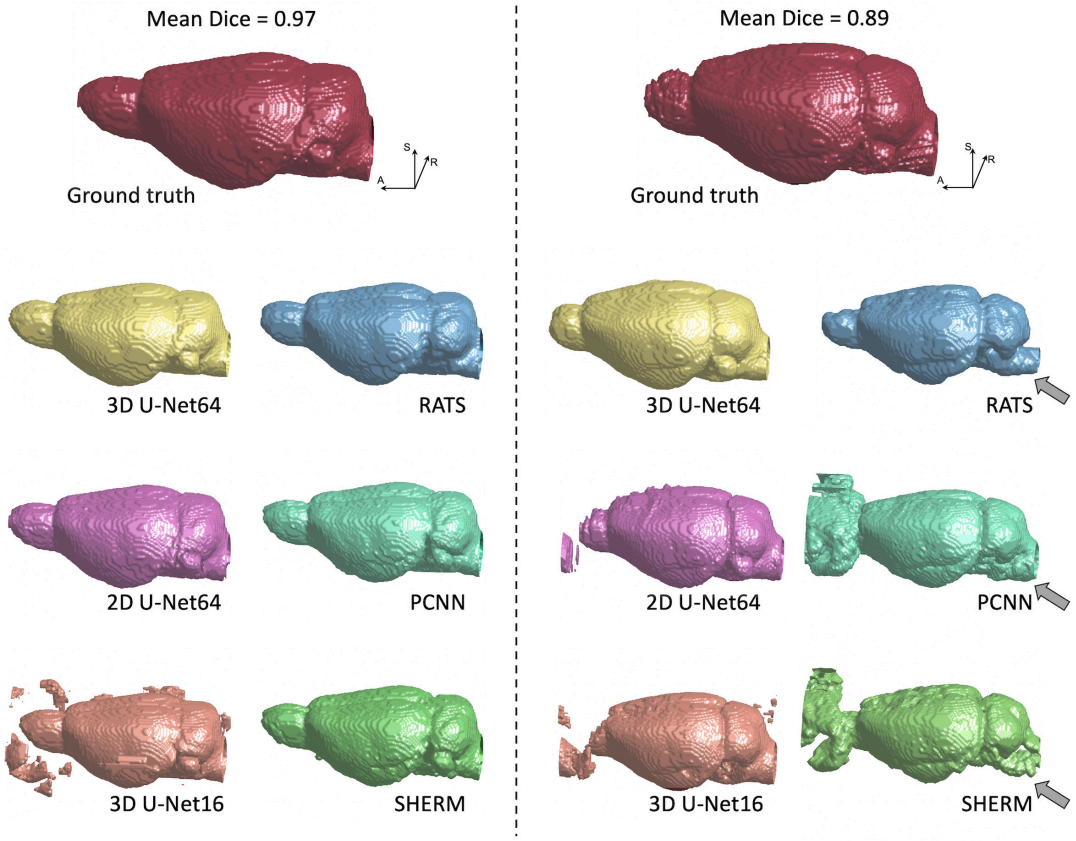


FIGURE 4 | 3D rendering of identified brain masks on the best and worst-case subjects for the T2w RARE rat dataset. Selection was based on the highest and lowest mean Dice score. Specifically, in the worst-case subject, 2D U-Net64, 3D U-Net16, and RATS missed the olfactory bulb, whereas PCNN and SHERM overestimated the olfactory bulb and incorporated surrounding frontal regions. Additionally, RATS, PCNN, and SHERM are missing significant portions of the cerebellum and brainstem (gray arrows). 3D U-Net64 and 3D U-Net32 produces excellent brain segmentation on both best and worst-case subjects.

There were no significant differences in each training sample selection across varied validation sample sizes in T2w RARE and T2*w EPI image segmentation, suggesting that a large number of model-validation subjects is perhaps unnecessary.

DISCUSSION

U-Net is a neural network that is mainly used for classification and localization (Ronneberger et al., 2015; Çiçek et al., 2016). In this study, we proposed a 3D U-Net architecture for automatic rat brain segmentation. Our results indicate that our proposed brain extraction framework based on 3D U-Net64 represents a robust method for the accurate and automatic extraction of rat brain tissue from MR images.

In this study where isotropic T2w RARE and T2* EPI data were segmented, RATS, PCNN, and SHERM still performed quickly and accurately (Dice > 0.8). 3D U-Net64 showed superior performance (average Dice = 0.99) over these methods and also outperformed our previously proposed 2D U-Net64 [Hsu et al., 2020; please also refer to Hsu et al. (2020) for detailed discussion about RATS, PCNN, and SHERM]. The 3D U-Net algorithm remains "parameter free" in the segmentation process, as all parameters are automatically learned from the data itself. It should be noted that the U-Net architecture generally requires longer processing times and needs a higher level of computational power for model-training. For single patch inference, 2D U-Net takes less time than 3D U-Net. However, the patch size of 3D U-Net64 is one more dimension than that of 2D U-Net64, so to obtain the whole volume prediction, 3D U-Net needs less inference time and less time overall.

3D U-Net is a variant of 2D U-Net where the inputs are 3D volume (Çiçek et al., 2016). It has the same encoder and decoder structure as in 2D U-Net. However, the encoder uses 3D convolution followed by 3D max-pooling to extract the features and the decoder uses 3D upsampling to reconstruct the annotated images (Çiçek et al., 2016). The key advantage of 3D U-Net is its ability to utilize interslice contextual information. The 3D U-Net16 model tends to have higher SEN but lower PPV than the 2D U-Net64 model, as it misclassifies more background areas as brain (i.e., has more false positives). In other words, 2D U-Net64 may achieve higher precision because it makes fewer false positive errors. Although 3D U-Net16 can extract 3D context from the additional dimension that 2D U-Net64 lacks, its

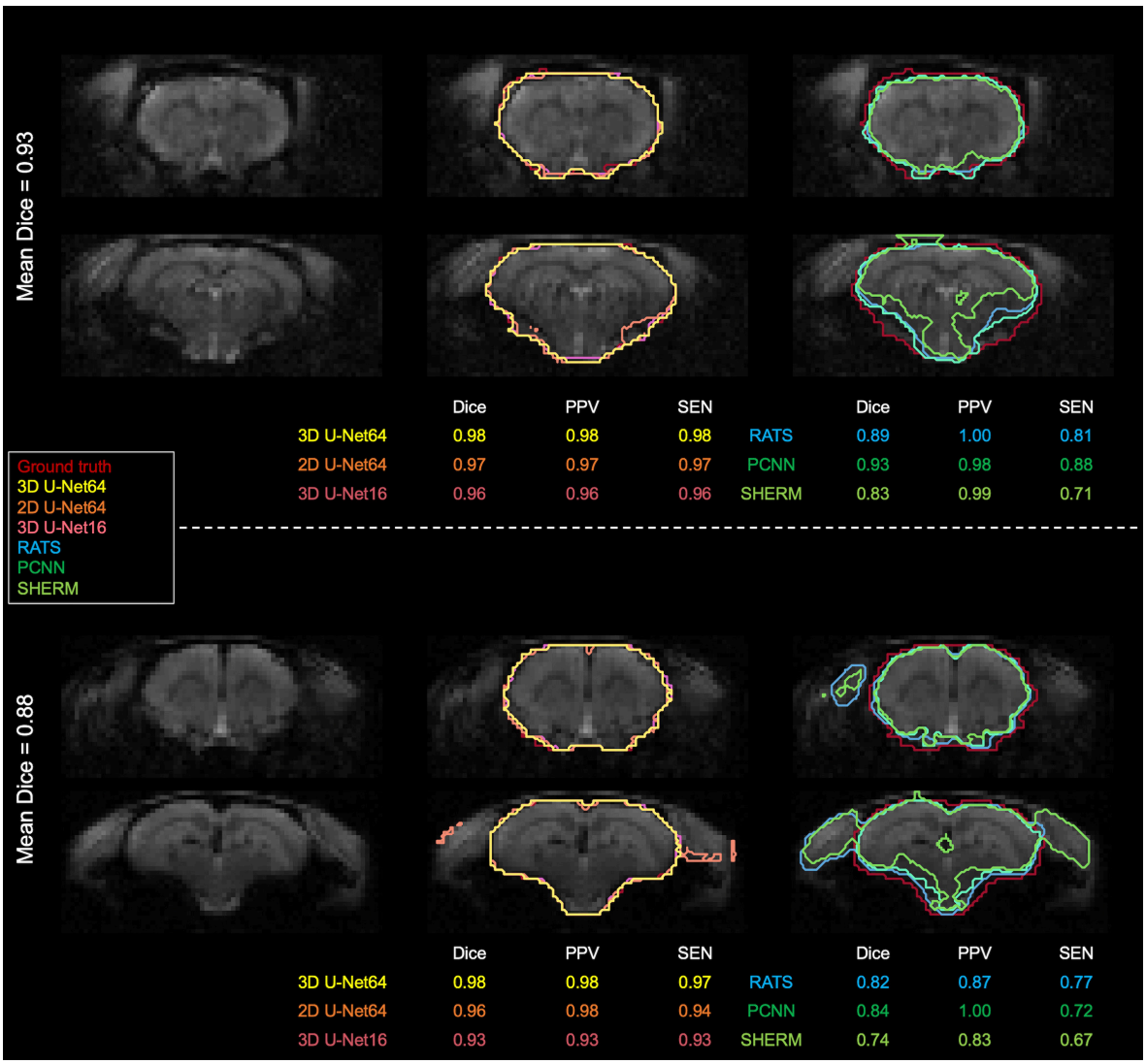


FIGURE 5 | Best and worst segmentation comparisons for T2*w EPI images. Selection was based on the highest and lowest mean Dice score (listed above the brain map) averaged over the six methods (3D U-Net64, 2D U-Net64, 3D U-Net16, RATS, PCNN, and SHERM). Posterior and inferior slices are more susceptible to error in RATS, PCNN, and SHERM, whereas all U-Net algorithms are more similar to the ground truth (all Dice > 0.90).

original 2D dimension resolution is also lost which means that the receptive field is decreased. Therefore, 2D U-Net64 achieves better performance than 3D U-Net16.

To illustrate the reliability of our proposed 3D U-Net64 architecture, we examined its accuracy under two conditions: conducting different Gaussian noise in the testing images and using a different number of rats from the original cohort in the training process. The results highlight the stability and robustness of 3D U-Net64 in segmenting the rat brain in both isotropic T2w RARE and T2*w EPI data. It is encouraging to observe excellent 3D U-Net64 performance (Dice > 0.95), even when the original image SNR is halved. Such performance is comparably higher than the conventional non-UNet-based methods using original SNR (RATS, PCNN, and SHERM). This demonstrates the noise-resistant capacity of 3D U-Net64 and suggests this approach may handle MRI data over a wide range of SNR quality. In addition, we also retrained the 3D U-Net64 model using a different number of subjects from the original cohort in the training process. This information is useful since the number of samples in a dataset are usually limited in rodent MRI studies. The Dice coefficient increases and stabilizes rapidly with >10 training subjects for both T2w RARE and T2*w EPI. This finding demonstrates the utility of 3D U-Net64 with a limited number of training datasets.

There are several limitations of the 3D U-Net architecture. First, deep learning is a data driven classification, so segmentation accuracy highly relies on the training dataset. Because we trained our 3D U-Net algorithm by using only T2w RARE and T2*w EPI images in rats, additional training and optimization will be needed to segment brain images with different contrast (e.g., T1-weighted images). Second, patch-based training could lose

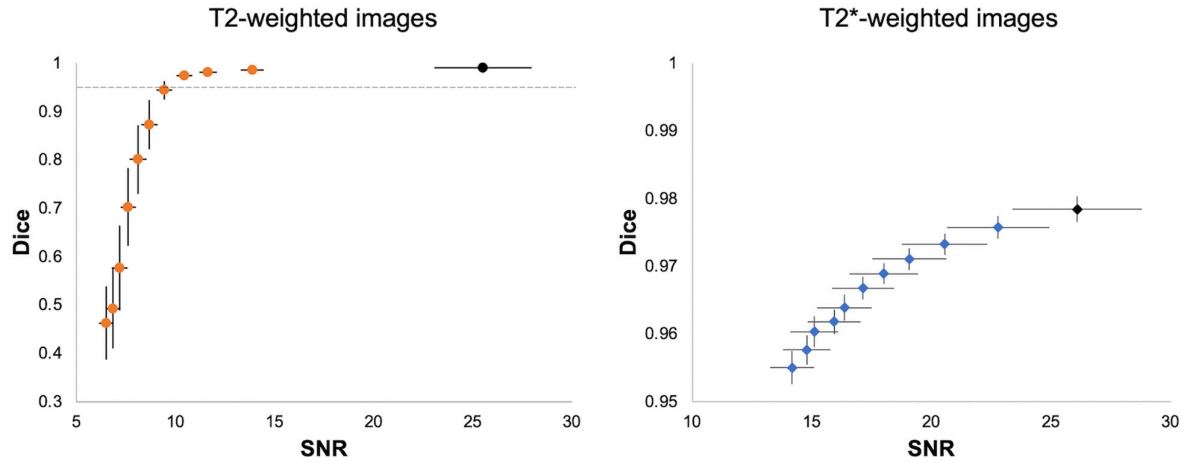


FIGURE 6 | Segmentation performance of 3D U-Net64 with different image SNR. For each T2w (*left*) and T2*w image (*right*), we added noise with random Gaussian distribution in the normalized testing images with variance from 5×10^{-5} to 5×10^{-4} and increments of 5×10^{-5} to investigate the segmentation performance of 3D U-Net64. Black dots indicate the averaged SNR and Dice from the original images without adding noise. The horizontal dot line in left panel indicates a Dice of 0.95. Error bar represents the standard error of Dice and SNR.

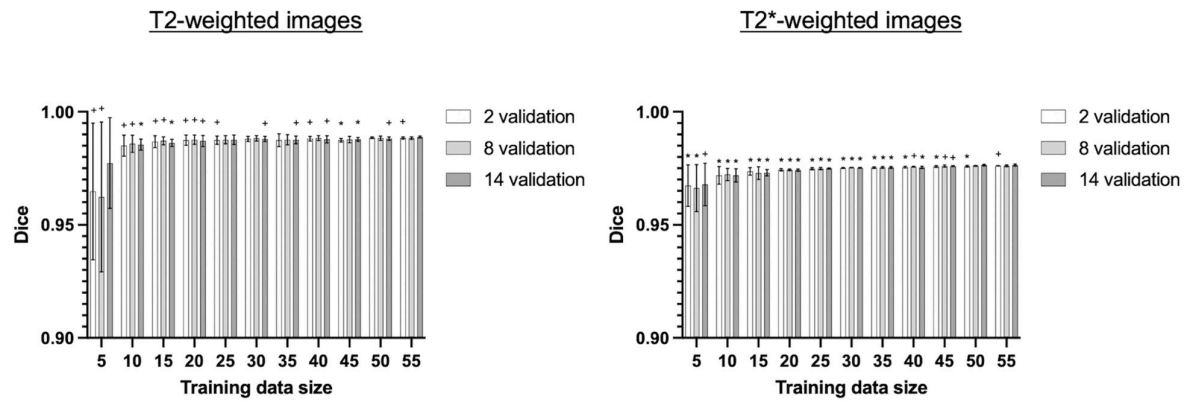


FIGURE 7 | Segmentation performance of 3D U-Net64 across different model-training and model-validating sample sizes. The 3D U-Net64 was trained from randomly selected subgroups. In the training process, we randomly selected 5–55 training subjects in increments of 5 subjects from the total 55 training dataset, and 2, 8, and 14 validation subjects from the total 14 validation dataset. The random selection was repeated 5 times to avoid bias. Statistical analyses compared Dice values under various conditions against the Dice values obtained from 55 training rats and 14 validation rats (one tailed paired *t*-test, $^{*}p < 0.05$, $^{+}0.05 < p < 0.1$). No significant differences were found between various model-validation sample selections within each model-training data selection for both T2w RARE and T2*w EPI data (repeated measurement ANOVA).

information/segmentation consistency or overfit the data if the patch size and number of training samples are imbalanced. The training sample imbalance may be alleviated by using randomly cropped patches to enhance sample diversity and prevent overfitting. Loss of information could be addressed by using a larger patch size. In this study, the image size of our whole rat brain T2w RARE data is 144 × 144 × 64 with resolution of 0.2 mm. Accordingly, we chose the patch size to be 64 × 64 × 64 (~2,097 mm^3), which could cover a significant portion of the rat brain (~2,000 mm^3). This patch size was limited by the slice dimension of the imaging matrix. Considering computation efficiency, we chose this patch size to also balance GPU demands for the users of this tool. By extracting overlapped patches with a 16 × 16 × 16 stride and using the averaged predictions for the final segmentation, we reduced segmentation inconsistency between neighboring patches and mitigated information loss. Our results support the use of a large patch size as we demonstrated superior segmentation performance at 64 × 64 × 64 patch in 3D U-Net than the 16 × 16 × 16 patch. Third, deep learning methods require substantial amounts of manually labeled data (Verbraeken et al., 2020), and their performance can be affected by similarities between the training dataset and the data entering the U-Net brain segmentation process. While more training datasets of varying quality may further improve our current 3D U-Net model, our validation analysis suggested reliable segmentation

performance can be reached with ≥10 training subjects. This indicates that the model can be established on a laboratory- or scanner-level basis. Next, our current 3D U-Net architecture uses a patch size of 64 × 64 × 64. As discussed above, this limits the testing image to an image matrix size of at least 64 × 64 × 64. Image resampling to a finer resolution is required if the image matrix size is smaller than 64 × 64 × 64. For a dataset at lower spatial resolution, a patch size of 16 × 16 × 16 as shown in this manuscript could be considered, although inferior performance may be expected. Alternatively, the source codes provided herein can be modified to adapt other patch sizes (such as 32 × 32 × 32) per the user's discretion. For datasets with highly anisotropic resolution (e.g., high in-plane resolution with very few slices), because our results show the 2D U-Net (64 × 64) significantly outperformed than the 3D U-Net (16 × 16 × 16), we suggest that the users consider 2D U-Net as described in Hsu et al. (2020). Finally, it should be noted that several advances in U-Net have been proposed, and thus the approach described herein may be improved further (e.g., nnU-Net, cascaded U-Net, U-Net++, double U-Net, and recurrent residual U-Net [R2U-Net]) (Alom et al., 2018; Zhou et al., 2018; Liu et al., 2019; Jha et al., 2020; Isensee et al., 2021). Additionally, several other deep learning-based methods, such as deep attention fully convolutional neural network (FCN) and Mask region-based convolutional neural network (R-CNN), have shown great promises for human medical image segmentation (Liu et al., 2018; Mishra et al., 2018; Chiao et al., 2019; Lei et al., 2020). Although one could fine-tune parameters and apply domain adaption strategies (Ghafoorian et al., 2017; Gholami et al., 2018) to attempt the use of human brain segmentation tools in rodents, no literature has documented success with this approach to the best of our knowledge, likely due to the differences in tissue geometry and MRI contrast features between human and rodent brains (Van Essen et al., 2013; Ma et al., 2018; Grandjean et al., 2020; Lee et al., 2021). Our study describes a step-by-step pipeline for extracting rat brain from isotropic MRI data. We focused primarily on 3D vs. 2D U-Net (Hsu et al., 2020) as well as comparisons against other established techniques developed specifically for rodent brain MRI (Chou et al., 2011; Oguz et al., 2014; Liu et al., 2020). Although additional comparisons with recently developed deep learning-based brain segmentation tools are beyond the scope of the present study, adapting those tools for rodent applications and establishing comprehensive evaluations should bring significant impact to the field. It is our hope that the work presented herein will facilitate such studies in the future.

CONCLUSION

We proposed a 3D U-Net model, an end-to-end deep learning-based segmentation algorithm for brain extraction of 3D high resolution volumetric brain MRI data. The method is fully automated and demonstrates accurate brain mask delineations for isotropic structural (T2w RARE) and functional (T2*w EPI) MRI data of the rat brain. The method was compared against other techniques commonly used for rodent brain extraction as well as 2D U-Net. 3D U-Net shows superior performance in qualitative metrics including Dice, Jaccard, PPV, SEN, CMD, and Hausdorff distance. We believe this tool will be useful to avoid parameter-selection bias and streamline pre-processing steps when analyzing high resolution 3D rat brain MRI data. The 3D U-Net brain extraction tool can be found at https://camri.org/dissemination/software/.

DATA AVAILABILITY STATEMENT

The CAMRI rats dataset is available at doi: 10.18112/openneuro.ds003646.v1.0.0. The 3D U-Net brain extraction tool can be found at https://camri.org/dissemination/software/. The 3D U-Net results of testing data are available at doi: 10.17632/gmcf67fw5h.1. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee at the University of North Carolina (UNC) at Chapel Hill.

AUTHOR CONTRIBUTIONS

L-MH and Y-YS designed the study. L-MH and SW implemented U-Net algorithm for rodents and validated the developed methods on various datasets. T-WW and S-HL provided ground-truth brain masks. LW provided data and helped to edit the manuscript. S-HL managed data and software dissemination and helped to design the study. L-MH, SW, and Y-YS wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Institute of Neurological Disorders and Stroke (R01NS091236 and R44NS105500), National Institute of Mental Health (R01MH126518, RF1MH117053, R01MH111429, and F32MH115439), National Institute on Alcohol Abuse and Alcoholism (P60AA011605 and U01AA020023), and National Institute of Child Health and Human Development (P50HD103573).

ACKNOWLEDGMENTS

We thank CAMRI members for insightful discussion on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.801008/full#supplementary-material

REFERENCES

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv* [preprint]. doi: 10.1109/NAECON.2018.8556686

Avants, B. B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTS). *Insight J.* 2, 1–35.

Babalola, K. O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., et al. (2009). An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage* 47, 1435–1447. doi: 10.1016/j.neuroimage.2009.05.029

Chiao, J.-Y., Chen, K.-Y., Liao, K. Y.-K., Hsieh, P.-H., Zhang, G., and Huang, T.-C. (2019). Detection and classification the breast tumors using mask R-CNN on sonograms. *Medicine* 98:e15200. doi: 10.1097/MD.0000000000015200

Chollet, F. (2015). *Keras Documentation [WWW Document]*. San Francisco: GitHub

Chou, N., Wu, J., Bai Bingren, J., Qiu, A., and Chuang, K.-H. (2011). Robust automatic rodent brain extraction using 3-D pulse-coupled neural networks (PCNN). *IEEE Trans. Image Process.* 20, 2554–2564. doi: 10.1109/TIP.2011.2126587

Çiçek, Ö, Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Lecture Notes in Computer Science*, eds S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells (Cham: Springer International Publishing), 424–432. doi: 10.1007/978-3-319-46723-8_49

Fatima, A., Shahid, A. R., Raza, B., Madni, T. M., and Janjua, U. I. (2020). State-of-the-Art Traditional to the Machine- and Deep-Learning-Based Skull Stripping Techniques, Models, and Algorithms. *J. Digit. Imaging* 33, 1443–1464. doi: 10.1007/s10278-020-00367-5

Feo, R., and Giove, F. (2019). Towards an efficient segmentation of small rodents brain: a short critical review. *J. Neurosci. Methods* 323, 82–89. doi: 10.1016/j.jneumeth.2019.05.003

Gaser, C., Schmidt, S., Metzler, M., Herrmann, K.-H., Krumbein, I., Reichenbach, J. R., et al. (2012). Deformation-based brain morphometry in rats. *Neuroimage* 63, 47–53. doi: 10.1016/j.neuroimage.2012.06.066

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., and Pesteie, M. (2017). "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Quebec City: Springer), 516–524.

Gholami, A., Subramanian, S., and Shenoy, V. (2018). "A novel domain adaptation framework for medical image segmentation," in *International MICCAI Brainlesion Workshop 2018*. (Cham: Springer) 289–298

Grandjean, J., Canella, C., Anckaerts, C., Ayrancı, G., Bougacha, S., Bienert, T., et al. (2020). Common functional networks in the mouse brain revealed by multi-centre resting-state fMRI analysis. *Neuroimage* 205:116278. doi: 10.1016/j.neuroimage.2019.116278

Hsu, L.-M., Wang, S., Ranadive, P., Ban, W., Chao, T.-H. H., Song, S., et al. (2020). Automatic Skull Stripping of Rat and Mouse Brain MRI Data Using U-Net. *Front. Neurosci.* 14:568614. doi: 10.3389/fnins.2020.568614

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18, 203–211. doi: 10.1038/s41592-020-01008-z

Jha, D., Riegler, M. A., and Johansen, D. (2020). "Doubleu-net: A deep convolutional neural network for medical image segmentation," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, (Piscataway: IEEE), 558–564.

Kingma, D. P., and Ba, J. L. (2015). "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations (ICLR)*, (Ithaca: arXiv.org), 13.

Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., et al. (2016). Deep MRI brain extraction: a 3D convolutional neural network for skull stripping. *Neuroimage* 129, 460–469. doi: 10.1016/j.neuroimage.2016.01.024

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing System*, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Red Hook: Curran Associates, Inc).

Lee, S.-H., Broadwater, M. A., Ban, W., Wang, T.-W. W., Kim, H.-J., Dumas, J. S., et al. (2021). An isotropic EPI database and analytical pipelines for rat brain resting-state fMRI. *Neuroimage* 243:118541. doi: 10.1016/j.neuroimage.2021.118541

Lei, Y., Dong, X., Tian, Z., Liu, Y., Tian, S., Wang, T., et al. (2020). CT prostate segmentation based on synthetic MRI-aided deep attention fully convolution network. *Med. Phys.* 47, 530–540. doi: 10.1002/mp.13933

Liu, H., Shen, X., Shang, F., Ge, F., and Wang, F. (2019). "CU-Net: Cascaded U-Net with Loss Weighted Sampling for Brain Tumor Segmentation," in *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy: 4th International Workshop, MBIA 2019, and 7th International Workshop, MFCA 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings, Lecture Notes in Computer Science*, eds D. Zhu, J. Yan, H. Huang, L. Shen, P. M. Thompson, and C.-F. Westin (Cham: Springer International Publishing), 102–111. doi: 10.1007/978-3-030-33226-6_12

Liu, M., Dong, J., Dong, X., and Yu, H. (2018). "Segmentation of lung nodule in CT images based on mask R-CNN," in *2018 9th International Conference on Awareness Science and Technology, iCAST*, (Piscataway: IEEE), 95–100.

Liu, Y., Unsal, H. S., Tao, Y., and Zhang, N. (2020). Automatic brain extraction for rodent MRI images. *Neuroinformatics* 18, 395–406. doi: 10.1007/s12021-020-09453-z

Lu, H., Scholl, C. A., Zuo, Y., Demny, S., Rea, W., Stein, E. A., et al. (2010). Registering and analyzing rat fMRI data in the stereotaxic framework by exploiting intrinsic anatomical features. *Magn. Reson. Imaging* 28, 146–152. doi: 10.1016/j.mri.2009.05.019

Ma, Z., Perez, P., Ma, Z., Liu, Y., Hamilton, C., Liang, Z., et al. (2018). Functional atlas of the awake rat brain: a neuroimaging study of rat brain specialization and integration. *Neuroimage* 170, 95–112. doi: 10.1016/j.neuroimage.2016.07.007

Mandino, F., Cerri, D. H., Garin, C. M., Straathof, M., van Tilborg, G. A. F., Chakravarty, M. M., et al. (2019). Animal functional magnetic resonance imaging: trends and path toward standardization. *Front. Neuroinform.* 13:78. doi: 10.3389/fninf.2019.00078

Martín, A., Paul, B., Jianmin, C., Zhifeng, C., and Andy, D. (2016). "Tensorflow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, (California: USENIX Association), 265–283.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, (Piscataway: IEEE), 565–571. doi: 10.1109/3DV.2016.79

Mishra, D., Chaudhury, S., Sarkar, M., and Soin, A. S. (2018). Ultrasound Image Segmentation: a Deeply Supervised Network with Attention to Boundaries. *IEEE Trans. Biomed. Eng.* 66, 1637–1648. doi: 10.1109/TBME.2018.2877577

Oguz, I., Zhang, H., Rumple, A., and Sonka, M. (2014). RATS: rapid Automatic Tissue Segmentation in rodent brain MRI. *J. Neurosci. Methods* 221, 175–182. doi: 10.1016/j.jneumeth.2013.09.021

Paxinos, G., and Watson, C. (2007). *The Rat Brain in Stereotaxic Coordinates: Hard Cover Edition*, 6th Edn. Amsterdam: Academic Press.

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), Lecture Notes in Computer Science*, eds N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Cham: Springer International Publishing), 234–241. doi: 10.1007/978-3-319-24574-4_28

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29, 1310–1320. doi: 10.1109/TMI.2010.2046908

Uhlich, M., Greiner, R., Hoehn, B., Woghiren, M., Diaz, I., Ivanova, T., et al. (2018). Improved Brain Tumor Segmentation via Registration-Based Brain Extraction. *Forecasting* 1, 59–69. doi: 10.3390/forecast1010005

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., et al. (2013). The WU-Minn Human Connectome Project: an overview. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041

Verbraeken, J., Wolting, M., Katzy, J., Kloppenburg, J., Verbelen, T., and Rellermeyer, J. S. (2020). A survey on distributed machine learning. *ACM Comput. Surv.* 53, 1–33. doi: 10.1145/3377454

Wang, S., Nie, D., Qu, L., Shao, Y., Lian, J., Wang, Q., et al. (2020). CT male pelvic organ segmentation via hybrid loss network with incomplete annotation. *IEEE Trans. Med. Imaging* 39, 2151–2162. doi: 10.1109/TMI.2020.2966389

Wang, X., Li, X.-H., Cho, J. W., Russ, B. E., Rajamani, N., Omelchenko, A., et al. (2021). U-net model for brain extraction: trained on humans for transfer to non-human primates. *Neuroimage* 235:118001. doi: 10.1016/j.neuroimage.2021.118001

Yogananda, C. G. B., Wagner, B. C., Murugesan, G. K., Madhuranthakam, A., and Maldjian, J. A. (2019). "A deep learning pipeline for automatic skull stripping and brain segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, (Piscataway: IEEE), 727–731. doi: 10.1109/ISBI.2019.8759465

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support*, ed. D. Stoyanov (Cham: Springer), 3–11. doi: 10.1007/978-3-030-00889-5_1

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

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

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fnins.2021.785595

Discriminative Analysis of Schizophrenia Patients Using Topological Properties of Structural and Functional Brain Networks: A Multimodal Magnetic Resonance Imaging Study

Jing Wang[1], Pengfei Ke[2,6], Jinyu Zang[2,6], Fengchun Wu[3,4] and Kai Wu[2,3,4,5,6,7,8,9]**

[1] School of Biomedical Engineering, Guangzhou Xinhua University, Guangzhou, China, [2] School of Biomedical Sciences and Engineering, South China University of Technology, Guangzhou, China, [3] The Affiliated Brain Hospital of Guangzhou Medical University, Guangzhou, China, [4] Guangdong Engineering Technology Research Center for Translational Medicine of Mental Disorders, Guangzhou, China, [5] Guangdong Engineering Technology Research Center for Diagnosis and Rehabilitation of Dementia, Guangzhou, China, [6] National Engineering Research Center for Tissue Restoration and Reconstruction, South China University of Technology, Guangzhou, China, [7] Key Laboratory of Biomedical Engineering of Guangdong Province, South China University of Technology, Guangzhou, China, [8] Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, Guangzhou, China, [9] Department of Nuclear Medicine and Radiology, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Thomas Bolton,
Advanced Telecommunications Research Institute International (ATR), Japan
Ling-Li Zeng,
National University of Defense Technology, China
Linling Li,
Shenzhen University, China

***Correspondence:**
Fengchun Wu
13580380071@163.com
Kai Wu
kaiwu@scut.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 29 September 2021
Accepted: 01 December 2021
Published: 11 January 2022

Citation:
Wang J, Ke P, Zang J, Wu F and Wu K (2022) Discriminative Analysis of Schizophrenia Patients Using Topological Properties of Structural and Functional Brain Networks: A Multimodal Magnetic Resonance Imaging Study. Front. Neurosci. 15:785595. doi: 10.3389/fnins.2021.785595

Interest in the application of machine learning (ML) techniques to multimodal magnetic resonance imaging (MRI) data for the diagnosis of schizophrenia (SZ) at the individual level is growing. However, a few studies have applied the features of structural and functional brain networks derived from multimodal MRI data to the discriminative analysis of SZ patients at different clinical stages. In this study, 205 normal controls (NCs), 61 first-episode drug-naive SZ (FESZ) patients, and 79 chronic SZ (CSZ) patients were recruited. We acquired their structural MRI, diffusion tensor imaging, and resting-state functional MRI data and constructed brain networks for each participant, including the gray matter network (GMN), white matter network (WMN), and functional brain network (FBN). We then calculated 3 nodal properties for each brain network, including degree centrality, nodal efficiency, and betweenness centrality. Two classifications (SZ vs. NC and FESZ vs. CSZ) were performed using five ML algorithms. We found that the SVM classifier with the input features of the combination of nodal properties of both the GMN and FBN achieved the best performance to discriminate SZ patients from NCs [accuracy, 81.2%; area under the receiver operating characteristic curve (AUC), 85.2%; $p < 0.05$]. Moreover, the SVM classifier with the input features of the combination of the nodal properties of both the GMN and WMN achieved the best performance to discriminate FESZ from CSZ patients (accuracy, 86.2%; AUC, 92.3%; $p < 0.05$). Furthermore, the brain areas in the subcortical/cerebellum network and the

frontoparietal network showed significant importance in both classifications. Together, our findings provide new insights to understand the neuropathology of SZ and further highlight the potential advantages of multimodal network properties for identifying SZ patients at different clinical stages.

Keywords: schizophrenia, brain networks, discriminative analysis, machine learning, multimodal MRI

INTRODUCTION

Schizophrenia (SZ) is a chronic psychiatric disease with hallucinations, delusions, and cognitive dysfunction (Lu et al., 2016). With the development of magnetic resonance imaging (MRI), the vast majority of studies have shown structural and functional brain abnormalities in SZ patients (Yamasue et al., 2004; Antonova et al., 2005; Kuroki, 2006; Pagsberg et al., 2007; Zhou et al., 2007). Several structural MRI studies have reported SZ patients show hippocampal volume reduction and bilateral thalamus volume reduction compared with normal controls (NCs) (Adriano et al., 2010, 2012), whereas the most reported functional alterations in SZ patients are located in thalamus, medial frontal gyrus, and superior temporal gyrus (Ren et al., 2013; Turner, 2013). Furthermore, previous studies have indicated that structural brain abnormalities are more widespread in chronic SZ (CSZ) patients than in first-episode drug-naive SZ (FESZ) patients, suggesting the potential impact of antipsychotic medication on structural brain abnormalities (Moncrieff and Leo, 2010; Torres et al., 2016). On the other hand, numerous functional MRI studies indicated that CSZ patients showed significant reductions in functional characteristics in brain regions involved in auditory, visual processing, and sensorimotor functions compared with FESZ patients (Wu et al., 2018). Therefore, it is critical to develop neuroimaging-based biomarkers for distinguishing the illness stages of SZ patients.

A number of studies based on the quantitative analysis of brain networks have reported that SZ patients show significantly decreased connectivity between a range of brain regions, particularly involving connections among the frontal lobe, temporal lobe, and insula compared with NCs (Bassett et al., 2008; Palaniyappan et al., 2015; Erdeniz et al., 2017). Previous studies of gray matter networks (GMNs) have shown that SZ patients exhibit reduced betweenness centrality (BC) in several regions and increased BC mainly in primary cortex and paralimbic cortex regions (Zhang et al., 2012). Numerous studies of white matter networks (WMNs) have demonstrated that a decreased clustering coefficient (Zhou et al., 2007; Collin et al., 2014), decreased global efficiency (van den Heuvel et al., 2013; Drakesmith et al., 2015; Griffa et al., 2015), and decreased node efficiency (NE) of the frontal lobe and limbic system were found in SZ patients (van den Heuvel et al., 2010; Wang et al., 2012; Sun et al., 2015). In addition, previous studies of functional brain networks (FBNs) indicated that the degree centrality (DC) in SZ patients was decreased in the bilateral putamen and increased in the left superior frontal gyrus (Chen et al., 2015). However, a few studies compared FESZ with CSZ patients based on the quantitative analysis of brain networks.

Recently, machine learning (ML) methods using neuroimaging data have been increasingly applied in the classification between SZ patients and normal controls (NCs), in which the classification accuracy varies from 0.65 to 0.95 (Arbabshirani et al., 2013; Talpalaru et al., 2019; Cao et al., 2020; Guo et al., 2020; Steardo et al., 2020). The majority of previous studies have mainly applied ML methods to a single neuroimaging modality, including structural MRI (sMRI) (Xiao et al., 2017; Oh et al., 2020), diffusion tensor imaging (DTI) (Ingalhalikar et al., 2010; Ardekani et al., 2011), resting-state functional MRI (rs-fMRI) (Koch et al., 2015; Skåtun et al., 2017; Cai et al., 2020), and electroencephalogram (EEG) (Ke et al., 2021). More recently, some studies have used multimodal MRI data to detect SZ at the level of the individual, and most measures are derived from multimodal MRI features, such as gray matter volume, regional homogeneity, amplitude of low-frequency fluctuation, and degree of centrality (Lu et al., 2018; Liang et al., 2019; Lei D. et al., 2020). Our previous study has applied multimodal MRI features to distinguish the illness stages of SZ patients based on ML methods and the findings have contributed to stage-specific biomarkers in diagnosis and interventions of SZ (Wu et al., 2018).

The objective of this study is to apply the features of both structural and functional brain networks derived from multimodal MRI data to the discriminative analysis of SZ patients at different clinical stages. We constructed three types of brain networks, including GMN, WMN, and FBN, for each participant derived from sMRI, DTI and rs-fMRI data. Three nodal properties of each brain networks were calculated, including BC, DC, and NE. The performance of two classifications (SZ vs. NC and FESZ vs. CSZ) was analyzed by using different combinations of nodal properties as the input features. Five ML algorithms were applied in both discriminative analyses, including the linear support vector machine (SVM) (Cortes, 1995), random forest algorithm (RF) (Breiman, 1996), logistic regression (LR) (Peng et al., 2002), linear discriminant analysis (LDA) (Fisher, 1936), and K-nearest neighbor classification (KNN) (Zhang, 2016).

MATERIALS AND METHODS

Subjects

A total of 61 FESZ patients, 79 CSZ patients, and 205 NCs were recruited for this study. The SZ patients were diagnosed by trained and experienced clinical psychiatrists using a structured clinical interview according to Diagnostic and Statistical Manual of Mental Disorders: Fourth Edition, Text Revision (DSM-IV-TR) (Structured Clinical Interview for DSM

Disorders [SCID]) criteria. The FESZ, CSZ, and NC groups were recruited from the Affiliated Brain Hospital of Guangzhou Medical University and the local community, respectively. All subjects were aged between 18 and 45 years, and their biological parents were Han Chinese. Before scanning, a clinical assessment was performed by psychiatrists using the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987). The subjects obtained a consensus score for each item on all three subscales (positive symptoms, negative symptoms, and general psychopathology) that was based on a seven-point scale indicating the severity of the symptom (van Tol et al., 2014). The inclusion criteria for all SZ patients were as follows: (1) a total score of at least 60 for the three PANSS subscales and (2) at least 3 positive symptom items on the PANSS with a score of at least 4. Additionally, FESZ patients were recruited for the first time when they were seeking help due to psychotic symptoms and did not take any antipsychotics. All CSZ patients were taking antipsychotics, and the course of the disease was greater than 2 years.

The exclusion criteria for all subjects included (1) any other psychiatric Axis I disorder meeting DSM-IV criteria, including schizoaffective disorders, intellectual disability, major depressive disorder, bipolar disorder, delirium, dementia, memory disorder, and other cognitive disorders; (2) mental disorder due to substance dependence, a seriously unstable somatic disease, definite diabetes, thyroid diseases, hypertension, or heart disease; (3) narrow angle glaucoma; (4) a history of epilepsy, except for febrile convulsions; (5) alcohol dependence meeting DSM-IV-TR criteria (excluding nicotine dependence); (6) having received electroconvulsive therapy within the past 6 months; (7) a contraindication for MRI; (8) medical resource neuroleptic malignant syndrome or serious tardive dyskinesia; (9) a serious suicide attempt or an irritative state; (10) noncompliant drug administration or a lack of legal guardians; or (11) lactating, pregnant, planning to become pregnant. In addition, NCs were excluded if they had a first- or second-degree relative with a psychiatric Axis I disorder according to DSM-IV criteria. Before enrollment, all subjects or their legal guardians provided written informed consent. These studies were performed according to the Declaration of Helsinki and approved by the Ethics Committees of the Affiliated Brain Hospital of Guangzhou Medical University.

Magnetic Resonance Imaging Data Acquisition

MRI images were acquired using a Philips 3T MR system (Philips, Achieva, Netherlands) located at the Affiliated Brain Hospital of Guangzhou Medical University. The participants were instructed to keep their eyes closed, to relax but not fall asleep, and to move as little as possible. The sMRI data were obtained using a sagittal three-dimensional gradient-echo T1-weighted sequence (256 × 256 × 188 matrix with 1 mm × 1 mm × 1 mm spatial resolution, repetition time (TR) = 8.2 ms, echo time (TE) = 3.8 ms, flip angle = 7°, field of view (FOV) = 256 mm × 256 mm). The DTI data were acquired using a single-shot echo-planar imaging-based sequence with the following parameters: slice thickness is 3 mm, no gap, 50 axial slices, TR = 6,000 ms; TE = 70 ms; flip angle = 90°; FOV = 256 mm × 256 mm; spatial resolution = 2 mm × 2 mm × 3 mm; 33 nonlinear diffusion weighting directions with b = 1000 s/mm^2 and one image without diffusion weighting (b = 0 s/mm^2). The rs-fMRI data were collected using an echo-planar imaging (EPI) sequence (64 × 64 × 36 matrix with 3.44 mm × 3.44 mm × 4 mm spatial resolution, TR = 2,000 ms, TE = 30 ms, flip angle = 90°, FOV = 220 mm × 220 mm).

Image Processing

All T1-weighted MRI data processing was performed using the SPM8 software package[1] (Institute of Neurology, University College London, United Kingdom). First, each T1-weighted MRI was segmented into three tissue maps, including gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), using the new segmentation algorithm from SPM8. Second, a customized, population-specific template was created from the segmented tissue maps using the DARTEL template-creation tool. Third, all GM maps were warped to the custom template space using its corresponding smooth, reversible deformation parameters. A modulation was applied by locally multiplying tissue values by the Jacobian determinants derived from the special normalization step (Good et al., 2001). Spatial smoothing was not performed to avoid inducing artifactual signal overlap among spatially adjacent regions (Kong et al., 2021).

Preprocessing of the DTI dataset was implemented using PANDA[2], which is a pipeline toolbox for diffusion MRI analysis (Cui et al., 2013). The procedure mainly included skull stripping, simple motion and eddy current correction, and diffusion tensor/parameter calculation. Eddy current was an important factor for image deformation, and affine transformation was used to register the DTI image to the T1 image, which can effectively reduce the influence of head movement and eddy current.

Rs-fMRI data were preprocessed using SPM8 (see text footnote 1) Institute of Neurology, University College London, United Kingdom) and DPABI (Yan et al., 2016). First, the first 10 volumes of each functional time series were discarded because the initial signal was unstable. Second, the remaining volumes were corrected for different signal acquisition times and realigned to the first volume to correct for head motion. Then, the nuisance signals (Rigid-body 6 motion parameters, the white matter signal, and the cerebrospinal fluid signal) were regressed out from the data. Subsequently, all functional volumes were normalized using EPI templates and resampled to 3-mm isotropic voxels. The resampled data were bandpass (0.01–0.08 Hz) filtered to reduce low-frequency drift and high-frequency physiological noise and spatially smoothed with a 4-mm FWHM Gaussian kernel. All participants in the study did not have excessive head motion (<2 mm or 2° during realignment and <2 mm for mean motion during framewise assessment).

Brain Network Construction

The single-subject GMN for T1-weighted MRI data was constructed based on the GM volume images. A GMN here

[1]http://www.fil.ion.ucl.ac.uk/spm

[2]http://www.nitrc.org/projects/panda/

included a collection of nodes and edges interconnecting the nodes. Here, the nodes represent brain regions, and edges represent interregional similarity in the distributions of regional GM volume. To define the network nodes, the GM volume image was parcellated into 268 regions of interest (ROIs) using a 268-node functional atlas (Shen et al., 2013). The GM volume value of each of the 268 ROIs was then calculated. We utilized a Kullback-Leibler (KL) divergence-based similarity (KLS) measure to estimate the network edges (Kong et al., 2014). Specifically, the probability density function of these 268 values was estimated using kernel density estimation (KDE) with bandwidths chosen automatically. The regional probability distribution function (PDF) and the KL divergence between each pair of 268 regions in their PDFs were then calculated. The network edges were defined as KLS with a consecutive sparsity threshold, S, ranging from $0.1 < S < 0.4$ (interval = 0.01). All of the following network analyses were performed at each of the threshold levels in this range. For each of the network metrics, the estimated values under the range of 0.1-0.4 were integrated with area under curve (Wang et al., 2016).

The WMN for DTI data was also parcellated into 268 ROIs. According to deterministic white matter fiber bundle tractography, the fiber number (FN) connected in the ROIs was taken as the connection weight, and then the FN-weighted connection matrix was calculated. The parameters of fiber tractography were set: the turning angle between adjacent voxels was less than 45 degrees or fractional anisotropy was greater than 0.2 (Gong et al., 2009). The white matter connection was calculated using PANDA (Cui et al., 2013). The network edges were defined as the number of white matter fibers with the threshold $FN > 2$.

The FBN for fMRI data was constructed based on the main regional time series by averaging voxelwise time series data. Pearson's correlation coefficient of the interregional time series was defined to measure the relation between network nodes. Then, each of the resulting correlation matrices was converted into a series of weighted networks with a sparsity threshold, S, ranging from $0.1 < S < 0.4$ (interval = 0.01). All of the following network analyses were performed at each of the threshold levels in this range. For each of the network metrics, the estimated values under the range of 0.1-0.4 were integrated with area under curve (Liu et al., 2017).

The nodal properties of each region, including DC, NE, and BC, were then calculated using the GRETNA toolbox (Wang et al., 2015).

Statistical Analysis

Between-group differences in age and years of education were analyzed by one-way analysis of variance (ANOVA) using SPSS 22.0 software. A *post hoc* analysis was performed using Scheffé's method. A χ^2 test was performed to determine sex differences. Statistical significance was set at $p < 0.05$.

Discriminative Analysis

In this study, we randomly split the set of participants into two groups, including a training dataset and an independent testing dataset, at a ratio of 4:1. In the training dataset, 10-fold cross-validation was performed. Each time, the ninefold data were used for training, and onefold data were used for validation. Each feature is normalized to between 0 and 1 on the training set, and the normalized parameters are applied to the validation set. Then, the performance of each resulting classifier was evaluated using the independent testing dataset. Using this method, we achieved unbiased estimates of every classifier.

The performance of two classifications (SZ patients vs. NC and FESZ patients vs. CSZ) was analyzed using 5 ML algorithms, including SVM, RF, LR, LDA, and KNN. Moreover, we used the method of recursive feature elimination (RFE) (Lin et al., 2017) to iteratively remove redundant features while preserving discriminative features (**Figure 1**). The analyses for both classification tasks were performed by using the in-house software NEURO-LEARN[3], which is a solution for collaborative pattern analysis of neuroimaging data (Lei B. et al., 2020).

From the results obtained by NEURO-LEARN, we obtained the weight of each input feature from the output of the classifier, and the absolute value of the feature weights can quantify the contribution of the features to the classifier. In this study, we calculated the feature contribution based on the results by the best classification performance. In addition, we divided 268 ROIs into eight subnetworks (Finn et al., 2015), including medial frontal, frontoparietal, default mode, subcortical/cerebellum, motor, visual I, visual II, and visual association (**Figure 2**). We selected the top 5% of combined features (BC, DC, and NE) ranked by the weights in the best classification to discuss subnetwork contributions.

To compare the performance of the 5 ML algorithms, we plotted the receiver operating characteristic (ROC) curves and calculated the area under the ROC curve (AUC). A permutation test was applied to explore whether the AUC obtained by the proposed model was significantly higher than the AUC of a random guess by randomly permuting the labels of the training data 1,000 times prior to the training step followed by the entire classification process. Based on probability distributions, it is possible to test the null hypothesis. The statistical significance was set at $p < 0.05$.

RESULTS

Clinical and Demographic Characteristics

The clinical and demographic characteristics of all subjects are shown in **Table 1**. There were significant differences in age and years of education were noted between NCs and CSZ patients ($p < 0.05$); there were also significant differences in age and years of education between NCs and FESZ patients ($p < 0.05$). Moreover, there was no significant difference in the positive, negative, general, and total PANSS scores between FESZ and CSZ patients.

[3]https://github.com/Raniac/NEURO-LEARN

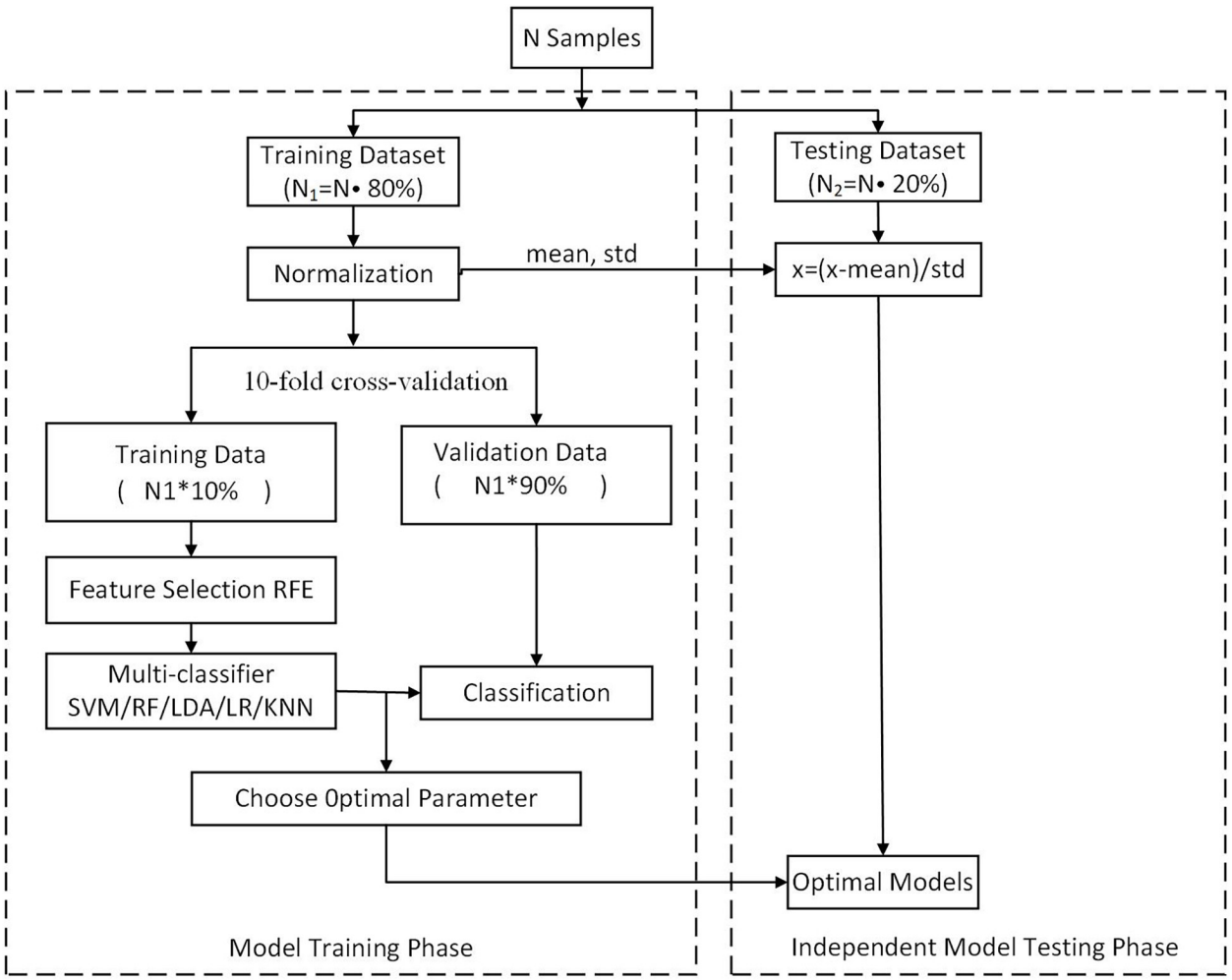


FIGURE 1 | Flowchart of the ML classification method.

Classification Performance

Two classifications were performed to distinguish SZ patients from NCs as well as FESZ patients from CSZ patients. The network properties of GMN, WMN, and FBN were used as the input features of the five ML algorithms. The results indicated that the SVM achieved the best performance compared to the other four ML algorithms based on RFE in both classifications (**Supplementary Tables 1, 2**).

Results of Classification Between Schizophrenia Patients and Normal Controls

When discriminating SZ patients from NCs, the classifier of SVM with the input features of the properties of GMN and FBN achieved the highest performance with an accuracy of 81.2% and AUC of 85.2% ($p < 0.05$), as shown in **Figure 3** and **Supplementary Table 1**.

We chose the top 5% of features ranked by their weights in the best classification (**Supplementary Table 3**). Meanwhile, we divided 268 ROIs into eight subnetworks and calculated the frequency of each subnetwork in the top 5% of features from the GMN and FBN. We found that the ROIs from the GMN were distributed in the subcortical/cerebellum network, frontoparietal network, motor network, medial frontal network, visual I network, visual II network, and visual association network. The ROIs from the FBN were distributed in the subcortical/cerebellum network, frontoparietal network, medial frontal network, visual II network, visual association network, default mode network, and motor network (**Figure 4**).

Overall, the ROIs were mainly distributed in the subcortical/cerebellum network (30%) and frontoparietal network (21.25%), and the proportion of the features from the GMN was greater than the proportion of the features from the FBN in the classification between SZ patients and NCs.

Results of Classification Between First-Episode Drug-Naive SZ and Chronic SZ Patients

When discriminating FESZ from CSZ patients, the SVM classifier with the input features of the properties of GMN and WMN achieved the highest performance with an accuracy of 86.2%

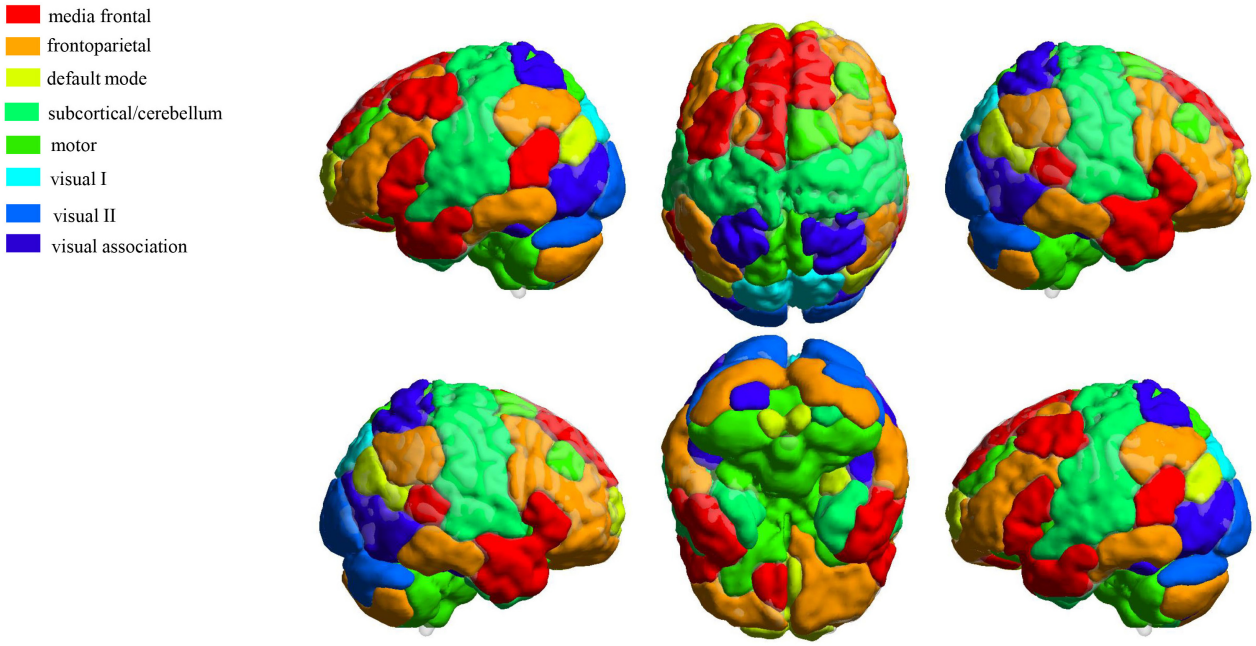


FIGURE 2 | Definition of eight functional networks using a 268-node functional atlas. The figure was generated by using the toolbox BrainNet Viewer (http://www.nitrc.org/projects/bnv/).

TABLE 1 | Demographic and clinical characteristics.

	FESZ patients	CSZ patients	NC	Statistic value	*p* value
Sex (F:M)	41:20	54:25	110:95	$\chi^2 = 3.53$	0.03
Age (years)	32.08 ± 7.42	33.21 ± 8.37	32.52 ± 8.40	$F = 5.39$	<0.05[a,b]
Education (years)	10.39 ± 3.25	11.97 ± 3.22	12.84 ± 2.83	$F = 21.33$	<0.05[a,b]
PANSS-PScore	24.02 ± 4.50	22.47 ± 5.70	–	$T = 1.74$	0.083
PANSS-NScore	21.64 ± 7.70	23.22 ± 7.29	–	$T = -1.24$	0.218
PANSS-GScore	40.31 ± 8.85	39.54 ± 10.18	–	$T = 0.47$	0.641
PANSS-TScore	85.97 ± 17.49	85.23 ± 19.44	–	$T = 0.23$	0.816

Values are represented as the mean ± standard deviation (SD). The comparisons of age and education among the three groups (FESZ, NC, and CSZ) were performed using a separate one-way ANOVA. Post hoc pairwise comparisons were then performed using a two-sample t-test. Statistical significance was set at $p < 0.05$. For the sex distribution among the three groups, the p value was obtained using the $\chi 2$ test.

[a]Post hoc paired comparisons showed significant group differences between CSZ vs. NC.

[b]Post hoc paired comparisons showed significant group differences between FESZ vs. NC.

CSZ, chronic schizophrenia; F, female; FESZ, first-episode drug-naive schizophrenia; GScore, general score; M, male; NC, normal control; NScore, negative syndrome score; PANSS, Positive and Negative Syndrome Scale; PScore, positive syndrome score; TScore, total syndrome score.

and AUC of 92.3% ($p < 0.05$), as shown in **Figure 5** and **Supplementary Table 2**.

We chose the top 5% of features ranked by the weights based on the best classification combination (**Supplementary Table 4**). We found that the ROIs from the GMN were distributed in the subcortical/cerebellum network, frontoparietal network, visual association network, motor network, medial frontal network, visual I network, and visual II network. The ROIs from the WMN were distributed in the subcortical/cerebellum network, frontoparietal network, motor network, visual II network, default mode network, visual I network, visual association network, and medial frontal network (**Figure 6**).

Overall, the ROIs were mainly distributed in the subcortical/cerebellum network (31.25%), frontoparietal network (13.75%) and motor network (15%), and the proportion of the features from the WMN was greater than the proportion of the features from the GMN in the discrimination FESZ from CSZ patients.

DISCUSSION

In this study, for the first time, we discriminated SZ patients at different clinical stages using the complex network properties of the GMN, WMN, and FBN. Our main findings were as follows: (1) the SVM algorithm achieved the best performance compared to the other four ML algorithms in both classifications (SZ vs. NC and FESZ vs. CSZ); (2) the

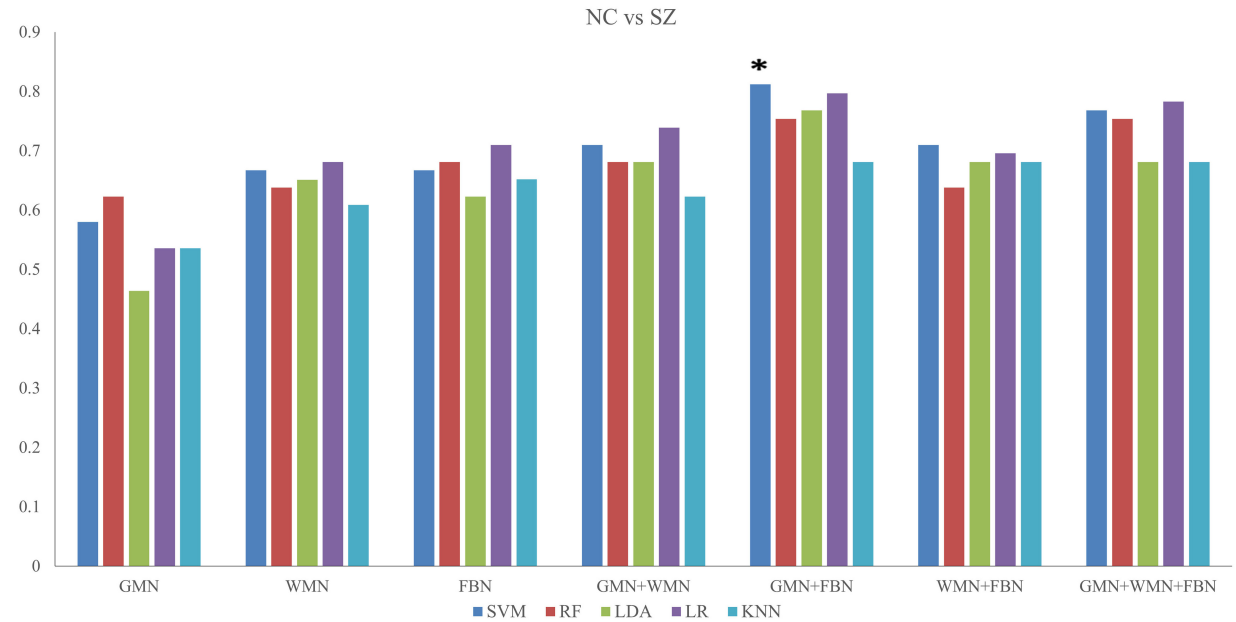


FIGURE 3 | Accuracy values of the classifications between SZ patients and NCs based on five ML methods. *The star indicates the highest values of best performance of the classifications between SZ patients and NCs.

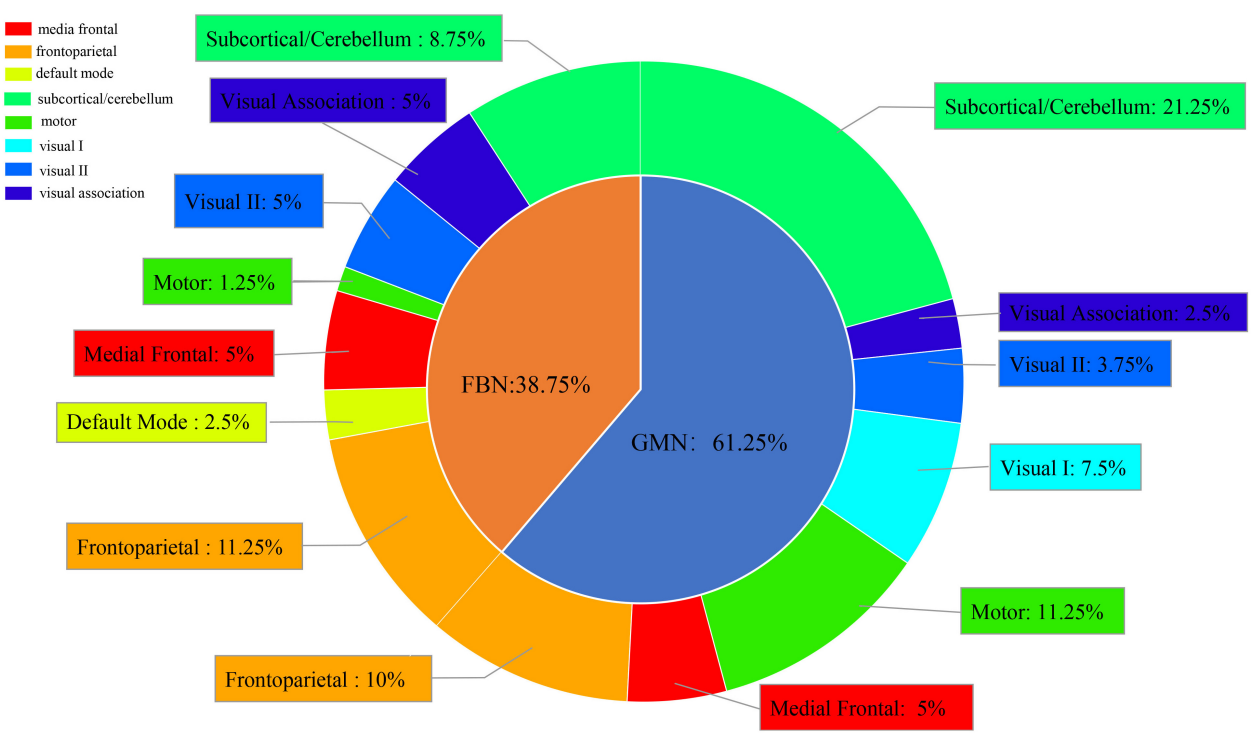


FIGURE 4 | Subnetwork distribution of ROIs from the FBN and GMN in the best classification between SZ patients and NCs.

classifier with the input features of GMN and FBN achieved the highest performance in the classification between SZ patients and NCs; the classifier with the input properties of GMN and WMN achieved the highest performance in the classification between FESZ and CSZ patients; and (3) the features of ROIs in both subcortical/cerebellum and frontoparietal networks showed significant importance in both classifications.



FIGURE 5 | Accuracy values of the classifications between FESZ and CSZ patients based on five ML methods. *The star indicates the highest values of best performance of the classifications between FESZ and CSZ patients.

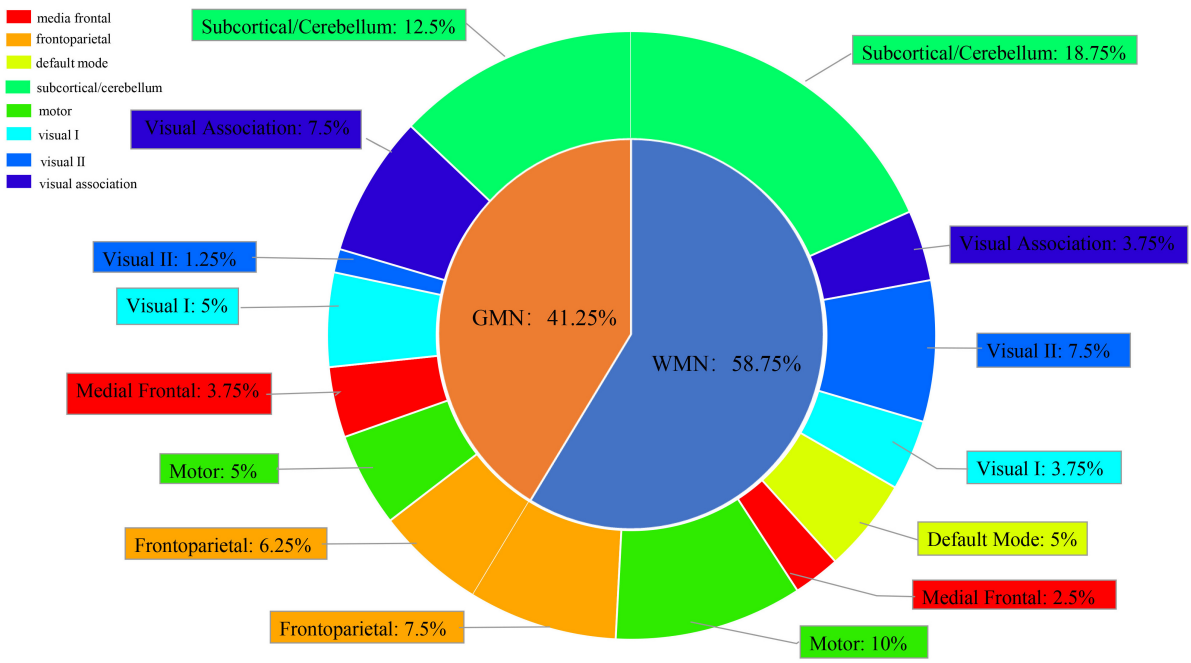


FIGURE 6 | Subnetwork distribution of ROIs from the GMN and WMN in the best classification between FESZ and CSZ patients.

The SVM algorithm, which determines a hyperplane that optimally distinguishes samples into two groups, has been widely used due to its reliable performance (Ding et al., 2015). In this study, our results indicated that the SVM achieved the best performance compared to the other four ML methods (including RF, LR, LDA, and KNN) in both classifications, which is consistent with previous studies of SZ (Chin et al., 2018; Rampisela and Rustam, 2018). A previous study discriminated SZ patients from NCs by analyzing multimodal brain imaging data with an SVM classifier and achieved good classification performance with a 91% accuracy and 100% prediction rate (Sui et al., 2014). Greenstein et al. (2012) applied the RF

method to discriminate SZ patients from NCs using 74 anatomic brain MRI subregions and achieved 73.7% accuracy. Winterburn et al. (2019) performed a critical appraisal of the accuracy of ML methodologies used in SZ patient and NC classifications by comparing three ML methods (including SVM, LR, and LDA), and the highest accuracy achieved was 73.5% using SVM. Similarly, Li et al. (2015) evaluated the classification performance of four classification methods (including LDA, KNN, SVM, and Gaussian process classifier) on SZ diagnosis and obtained a maximum accuracy of 85.83% using the SVM. A recent study reviewed five traditional ML algorithms (including SVM, RF, KNN, gradient boosting machine and naive Bayes) frequently used for mental health and indicated that the advantage of SVM was working relatively accurately in general in practice (Cho et al., 2019). Our results confirmed the effectiveness of the SVM algorithm on SZ classification with the features of structural and functional brain networks derived from multimodal MRI data.

The classifier with the input features of the combination of the nodal properties of both GMN and FBN achieved the best performance (accuracy of 81.2%) compared to the classifiers using the input features of a single brain network when discriminating SZ from NC. Considerable evidence indicates that SZ is associated with structural as well as functional brain abnormalities (Oh et al., 2017; Li et al., 2020). A recent study revealed that the use of combined structural and functional measures allows the highest accuracy of classification to detect SZ (Lei D. et al., 2020). However, the input features of previous SZ classification studies were derived from structural and functional neuroimaging features (de Filippis et al., 2019; Shi et al., 2021). Our results based on brain network properties further highlight the potential advantages of multimodal features for distinguishing SZ patients from NCs.

Importantly, we also found that the classifier using the nodal properties of both the GMN and WMN achieved the best performance (accuracy of 86.2%) when discriminating FESZ from CSZ. The WMN from DTI data played a key role in predicting individuals with FESZ, which is consistent with previous studies (Lee et al., 2013; Samartzis et al., 2014). Friedman et al. (2008) indicated that fractional anisotropy of the white matter in the forceps minor or the genu of the corpus callosum showed striking differences between FESZ and CSZ patients based on DTI findings. White matter abnormalities in SZ as estimated by DTI appear to be present in the early stage of the disorder, most likely reflecting the developmental stage (Kyriakopoulos and Frangou, 2009; Liang et al., 2019). The main objective of our study is to analyze the importance of complex brain networks derived from multi-modal MRI data in the classification of SZ patients at different clinical stages. Our results showed that GMN and FBN played a more important role in discriminating SZ from NCs, whereas GMN and WMN played a more important role in discriminating FESZ from CSZ. Together, these findings may provide new insights into the neuropathology of SZ at various clinical stages.

In this study, our results demonstrated that the ROIs in the subcortical/cerebellar network and frontoparietal network showed significant importance in SZ classification. A large number of previous studies have indicated that SZ patients showed significantly altered brain connectivity involved in the frontoparietal network, cerebellum network and visual network (Wang et al., 2018; de Filippis et al., 2019). Interestingly, the frontoparietal network and the cerebellum network are among the most commonly implicated networks in SZ patients (Watanabe et al., 2014). The frontoparietal network, which has multiple important hubs in the prefrontal cortex, is involved in executive processing and cognitive control and has been shown to exhibit abnormal activation and connectivity in SZ patients (Minzenberg et al., 2009; Cole et al., 2013; Tu et al., 2013). The cerebellum network, which is featured in the influential 'cognitive dysmetria' hypothesis of SZ patients (Andreasen et al., 1998), has been found to be the most important network in the functional connectivity of SZ patients according to the extended maximal information coefficient (Su et al., 2013). Moreover, previous studies based on deep learning algorithms have discovered that cortical-striatal-cerebellar functional connectivity features exhibit great weights in the classification of SZ (Zeng et al., 2018; Qureshi et al., 2019). We also found that the ROIs in these two networks showed key importance in discriminating FESZ and CSZ patients. A meta-analysis of 25 articles based on cerebellar structural and functional abnormalities in FESZ patients hypothesized that the changes in both structural and functional aspects might reflect a common pathophysiology of the cerebellum in SZ patients (Ding et al., 2019). Lui et al. (2010) reported that 6 weeks of antipsychotic treatment increased brain synchronous activity in the frontal and parietal regions in FESZ patients, which might contribute to the importance of the frontoparietal network in discriminating FESZ and CSZ patients. In addition, the ROIs in the motor network also showed significant importance in discriminating FESZ patients from CSZ patients but not in SZ patients from NCs. It was reported that the functional connectivity in the sensory-motor network could reflect the hypothesized "neurotoxic effect" of FESZ patients (Zhang et al., 2019), demonstrating that the motor network deserves more attention in the search for neuroimaging markers for evaluating neural impairment in SZ.

Interestingly, we found that the BC of network topological features has more importance than DC and NE to the classifier for both contrast groups. The BC is a ratio of the number of all shortest paths between any two nodes in the network that travel through an index node, providing an indication of how topologically central the role of node is in overall network communication (van den Heuvel and Fornito, 2014). Previous brain network studies have indicated that BC in some brain regions is decreased (van den Heuvel and Hulshoff Pol, 2010; Zhang et al., 2012). Our study may further confirm the importance of the BC of network topological features in identifying SZ.

LIMITATIONS OF THE METHODOLOGY

Possible study limitations should be considered. First, only three nodal features (BC, DC, and NC) were combined in the present study. Additional network measures, such as clustering coefficient, path length and global efficiency, can be used and may improve classification performance. Second, we compared

the performance of five ML methods using only RFE as the dimensionality reduction algorithm. Previous studies of SZ patients have shown different classification performances using different dimensionality reduction algorithms (Li et al., 2015; Winterburn et al., 2019). Third, we have used different brain atlases in the discriminative analysis of SZ patients based on neuroimaging features and revealed that the 268-node functional atlas outperformed the other two brain atlases (Zang et al., 2021). However, it is necessary to compare the effects among different combinations of dimensionality reduction algorithms and brain atlases from the perspective of brain network properties in future studies. Fourth, the classification performance did not always benefit from more modalities in either classification. Previous studies have also shown that including more features does not necessarily yield positive effects due to feature redundancy. However, a recent study proposed a safe classifier that could address this issue to some extent (Hou et al., 2019). It is important to study whether safe classification would affect the discriminative results based on multimodal brain networks in future studies.

CONCLUSION

We constructed GMNs, WMNs and FBNs from sMRI, DTI and fMRI data, respectively and then discriminated SZ patients at different clinical stages using different combinations of nodal properties based on five ML methods. Our results indicated the best performance of the SVM algorithm in SZ classifications and highlighted the potential advantages of multimodal network properties in identifying SZ patients at different stages. Furthermore, the ROIs in the subcortical/cerebellum network and frontoparietal network showed significant importance in both classifications (NC vs. SZ, FESZ vs. CSZ). Our findings may bring new insights into the understanding of the neuropathology of SZ from the perspective of network properties.

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 Ethics Committees of the Affiliated Brain Hospital of Guangzhou Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to data analysis, drafted and critically revised the manuscript, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

FUNDING

This work was supported by the National Key Research and Development Program of China (2020YFC2004300, 2020YFC2004301, 2019YFC0118800, 2019YFC0118802, 2019YFC0118804, 2019YFC0118805, 2021YFC2009400, and 2021YFC2009404), National Natural Science Foundation of China (31771074, 81802230, and 72174082), Guangdong Basic and Applied Basic Research Foundation Outstanding Youth Project (2021B1515020064), Key Research and Development Program of Guangdong (2018B030335001, 2020B0101130020, and 2020B0404010002), Science and Technology Program of Guangzhou (201807010064, 201803010100, 201903010032, and 202103000032), Key Laboratory Program of Guangdong Provincial Education Department (2020KSYS001), Open Fund of the Guangdong Provincial Key Laboratory of Physical Activity and Health Promotion (2021B1212040014),and Science Foundation of Guangzhou Xinhua University (2018YB002 and 2020YQYJ05).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.785595/full#supplementary-material

REFERENCES

Adriano, F., Caltagirone, C., and Spalletta, G. (2012). Hippocampal volume reduction in first-episode and chronic schizophrenia. *Neuroscientist* 18, 180–200. doi: 10.1177/1073858410395147

Adriano, F., Spoletini, I., Caltagirone, C., and Spalletta, G. (2010). Updated meta-analyses reveal thalamus volume reduction in patients with first-episode and chronic schizophrenia. *Schizophr. Res.* 123, 1–14. doi: 10.1016/j.schres.2010.07.007

Andreasen, N. C., Paradiso, S., and O'Leary, D. S. (1998). "Cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry? *Schizophr. Bull.* 24, 203–218. doi: 10.1093/oxfordjournals.schbul.a033321

Antonova, E., Kumari, V., Morris, R., Halari, R., Anilkumar, A., Mehrotra, R., et al. (2005). The relationship of structural alterations to cognitive deficits in schizophrenia: a voxel-based morphometry study. *Biol. Psychiatry* 58, 457–467. doi: 10.1016/j.biopsych.2005.04.036

Arbabshirani, M. R., Kiehl, K. A., Pearlson, G. D., and Calhoun, V. D. (2013). Classification of schizophrenia patients based on resting-state functional network connectivity. *Front. Neurosci.* 7:133. doi: 10.3389/fnins.2013.00133

Ardekani, B. A., Tabesh, A., Sevy, S., Robinson, D. G., Bilder, R. M., and Szeszko, P. R. (2011). Diffusion tensor imaging reliably differentiates patients with schizophrenia from healthy volunteers. *Hum. Brain Mapp.* 32, 1–9. doi: 10.1002/hbm.20995

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical

networks in health and schizophrenia. *J. Neurosci.* 28, 9239–9248. doi: 10.1523/JNEUROSCI.1929-08.2008

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 2, 123–140.

Cai, X. L., Xie, D. J., Madsen, K. H., Wang, Y. M., Bögemann, S. A., Cheung, E. F. C., et al. (2020). Generalizability of machine learning for classification of schizophrenia based on resting-state functional MRI data. *Hum. Brain Mapp.* 41, 172–184. doi: 10.1002/hbm.24797

Cao, B., Cho, R. Y., Chen, D., Xiu, M., Wang, L., Soares, J. C., et al. (2020). Treatment response prediction and individualized identification of first-episode drug-naïve schizophrenia using brain functional connectivity. *Mol. Psychiatry* 25, 906–913. doi: 10.1038/s41380-018-0106-5

Chen, C., Wang, H. L., Wu, S. H., Huang, H., Zou, J. L., Chen, J., et al. (2015). Abnormal degree centrality of bilateral putamen and left superior frontal gyrus in schizophrenia with auditory hallucinations: a resting-state functional magnetic resonance imaging study. *Chin. Med. J. (Engl.)* 128, 3178–3184. doi: 10.4103/0366-6999.170269

Chin, R., You, A. X., Meng, F., Zhou, J., and Sim, K. (2018). Recognition of schizophrenia with regularized support vector machine and sequential region of interest selection using structural magnetic resonance imaging. *Sci. Rep.* 8:13858. doi: 10.1038/s41598-018-32290-9

Cho, G., Yim, J., Choi, Y., Ko, J., and Lee, S. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry Investig.* 16, 262–269. doi: 10.30773/pi.2018.12.21.2

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., and Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355. doi: 10.1038/nn.3470

Collin, G., Kahn, R. S., de Reus, M. A., Cahn, W., and van den Heuvel, M. P. (2014). Impaired rich club connectivity in unaffected siblings of schizophrenia patients. *Schizophr. Bull.* 40, 438–448. doi: 10.1093/schbul/sbt162

Cortes, C. V. V. (1995). Support-vectorNetworks. *Mach. Learn.* 20, 273–297.

Cui, Z., Zhong, S., Xu, P., He, Y., and Gong, G. (2013). PANDA: a pipeline toolbox for analyzing brain diffusion images. *Front. Hum. Neurosci.* 7:42. doi: 10.3389/fnhum.2013.00042

de Filippis, R., Carbone, E. A., Gaetano, R., Bruni, A., Pugliese, V., Segura-Garcia, C., et al. (2019). Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review. *Neuropsychiatr. Dis. Treat.* 15, 1605–1627. doi: 10.2147/NDT.S202418

Ding, X., Yang, Y., Stein, E. A., and Ross, T. J. (2015). Multivariate classification of smokers and nonsmokers using SVM-RFE on structural MRI images. *Hum. Brain Mapp.* 36, 4869–4879. doi: 10.1002/hbm.22956

Ding, Y., Ou, Y., Pan, P., Shan, X., Chen, J., Liu, F., et al. (2019). Cerebellar structural and functional abnormalities in first-episode and drug-naive patients with schizophrenia: a meta-analysis. *Psychiatry Res. Neuroimaging* 283, 24–33. doi: 10.1016/j.pscychresns.2018.11.009

Drakesmith, M., Caeyenberghs, K., Dutt, A., Zammit, S., Evans, C. J., Reichenberg, A., et al. (2015). Schizophrenia-like topological changes in the structural connectome of individuals with subclinical psychotic experiences. *Hum. Brain Mapp.* 36, 2629–2643. doi: 10.1002/hbm.22796

Erdeniz, B., Serin, E., İbadi, Y., and Taş, C. (2017). Decreased functional connectivity in schizophrenia: the relationship between social functioning, social cognition and graph theoretical network measures. *Psychiatry Res. Neuroimaging* 270, 22–31. doi: 10.1016/j.pscychresns.2017.09.011

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., et al. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat.e Neurosci.* 18, 1664–1671. doi: 10.1038/nn.4135

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugen.* 2, 179–188.

Friedman, J. I., Tang, C., Carpenter, D., Buchsbaum, M., Schmeidler, J., Flanagan, L., et al. (2008). Diffusion tensor imaging findings in first-episode and chronic schizophrenia patients. *Am. J. Psychiatry* 165, 1024–1032. doi: 10.1176/appi.ajp.2008.07101640

Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., et al. (2009). Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb. Cortex* 19, 524–536. doi: 10.1093/cercor/bhn102

Good, C. D., Johnsrude, I., Ashburner, J., Henson, R. N. A., Friston, K. J., and Frackowiak, R. S. J. (2001). Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *NeuroImage* 14, 685–700. doi: 10.1006/nimg.2001.0857

Greenstein, D., Malley, J. D., Weisinger, B., Clasen, L., and Gogtay, N. (2012). Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front. Psychiatry* 3:53. doi: 10.3389/fpsyt.2012.00053

Griffa, A., Baumann, P. S., Ferrari, C., Do, K. Q., Conus, P., Thiran, J., et al. (2015). Characterizing the connectome in schizophrenia with diffusion spectrum imaging. *Hum. Brain Mapp.* 36, 354–366. doi: 10.1002/hbm.22633

Guo, Y., Qiu, J., and Lu, W. (2020). Support vector machine-based schizophrenia classification using morphological information from amygdaloid and hippocampal subregions. *Brain Sci.* 10:562. doi: 10.3390/brainsci10080562

Hou, C., Zeng, L., and Hu, D. (2019). Safe classification with augmented features. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 2176–2192. doi: 10.1109/TPAMI.2018.2849378

Ingalhalikar, M., Kanterakis, S., Gur, R., Roberts, T. P., and Verma, R. (2010). DTI based diagnostic prediction of a disease via pattern classification. *Med. Image Comput. Comput. Assist. Interv.* 13(Pt 1), 558–565. doi: 10.1007/978-3-642-15705-9_68

Kay, S. R., Fiszbein, A., and Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276.

Ke, P., Xiong, D., Li, J., Pan, Z., Zhou, J., Li, S., et al. (2021). An integrated machine learning framework for a discriminative analysis of schizophrenia using multi-biological data. *Sci. Rep.* 11:14636. doi: 10.1038/s41598-021-94007-9

Koch, S. P., Hägele, C., Haynes, J., Heinz, A., Schlagenhauf, F., and Sterzer, P. (2015). Diagnostic classification of schizophrenia patients on the basis of regional reward-related fMRI signal patterns. *PLoS One* 10:e119089. doi: 10.1371/journal.pone.0119089

Kong, L., Huang, Y., Lei, B., Ke, P., Li, H., Zhou, J., et al. (2021). Divergent alterations of structural–functional connectivity couplings in first-episode and chronic schizophrenia patients. *Neuroscience* 460, 1–12. doi: 10.1016/j.neuroscience.2021.02.008

Kong, X., Wang, X., Huang, L., Pu, Y., Yang, Z., Dang, X., et al. (2014). Measuring individual morphological relationship of cortical regions. *J. Neurosci. Methods* 237, 103–107. doi: 10.1016/j.jneumeth.2014.09.003

Kuroki, N. (2006). Middle and inferior temporal gyrus gray matter volume abnormalities in first-episode schizophrenia: an MRI study. *Am. J. Psychiatry* 163:2103. doi: 10.1176/ajp.2006.163.12.2103

Kyriakopoulos, M., and Frangou, S. (2009). Recent diffusion tensor imaging findings in early stages of schizophrenia. *Curr. Opin. Psychiatry* 22, 168–176. doi: 10.1097/YCO.0b013e328325aa23

Lee, S., Kubicki, M., Asami, T., Seidman, L. J., Goldstein, J. M., Mesholam-Gately, R. I., et al. (2013). Extensive white matter abnormalities in patients with first-episode schizophrenia: A diffusion tensor imaging (DTI) study. *Schizophr. Res.* 143, 231–238.

Lei, B., Wu, F., Zhou, J., Xiong, D., Wang, K., Kong, L., et al. (2020). NEURO-LEARN: a solution for collaborative pattern analysis of neuroimaging data. *Neuroinformatics* 19, 79–91. doi: 10.1007/s12021-020-09468-6

Lei, D., Pinaya, W. H. L., Young, J., Amelsvoort, T., Marcelis, M., Donohoe, G., et al. (2020). Integrating machining learning and multimodal neuroimaging to detect schizophrenia at the level of the individual. *Hum. Brain Mapp.* 41, 1119–1135. doi: 10.1002/hbm.24863

Li, A., Zalesky, A., Yue, W., Howes, O., Yan, H., Liu, Y., et al. (2020). A neuroimaging biomarker for striatal dysfunction in schizophrenia. *Nat. Med.* 26, 558–565. doi: 10.1038/s41591-020-0793-8

Li, Z., Wang, Y., Quan, W., Wu, T., and Lv, B. (2015). Evaluation of different classification methods for the diagnosis of schizophrenia based on functional near-infrared spectroscopy. *J. Neurosci. Methods* 241, 101–110. doi: 10.1016/j.jneumeth.2014.12.020

Liang, S., Li, Y., Zhang, Z., Kong, X., Wang, Q., Deng, W., et al. (2019). Classification of first-episode schizophrenia using multimodal brain features: a combined structural and diffusion imaging study. *Schizophr. Bull.* 45, 591–599. doi: 10.1093/schbul/sby091

Lin, X., Li, C., Zhang, Y., Su, B., Fan, M., and Wei, H. (2017). Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics. *Molecules* 23:52. doi: 10.3390/molecules23010052

Liu, Y., Wang, H., Duan, Y., Huang, J., Ren, Z., Ye, J., et al. (2017). Functional brain network alterations in clinically isolated syndrome and multiple sclerosis: a graph-based connectome study. *Radiology* 282, 534–541. doi: 10.1148/radiol.2016152843

Lu, X. B., Zhang, Y., Yang, D. Y., Yang, Y. Z., Wu, F. C., Ning, Y. P., et al. (2018). Analysis of first-episode and chronic schizophrenia using multi-modal magnetic resonance imaging. *Eur. Rev. Med. Pharmacol. Sci.* 22, 6422–6435. doi: 10.26355/eurrev_201810_16055

Lu, X., Yang, Y., Wu, F., Gao, M., Xu, Y., Zhang, Y., et al. (2016). Discriminative analysis of schizophrenia using support vector machine and recursive feature elimination on structural MRI images. *Medicine* 95:e3973.

Lui, S., Li, T., Deng, W., Jiang, L., Wu, Q., Tang, H., et al. (2010). Short-term effects of antipsychotic treatment on cerebral function in drug-naive first-episode schizophrenia revealed by "resting state" functional magnetic resonance imaging. *Arch. Gen. Psychiatry* 67:783. doi: 10.1001/archgenpsychiatry.2010.84

Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S., and Glahn, D. C. (2009). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Arch. Gen. Psychiatry* 66:811. doi: 10.1001/archgenpsychiatry.2009.91

Moncrieff, J., and Leo, J. (2010). A systematic review of the effects of antipsychotic drugs on brain volume. *Psychol. Med.* 40, 1409–1422. doi: 10.1017/S0033291709992297

Oh, J., Chun, J., Kim, E., Park, H., Lee, B., and Kim, J. (2017). Aberrant neural networks for the recognition memory of socially relevant information in patients with schizophrenia. *Brain Behav.* 7:e602. doi: 10.1002/brb3.602

Oh, J., Oh, B., Lee, K., Chae, J., and Yun, K. (2020). Identifying schizophrenia using structural MRI With a deep learning algorithm. *Front. Psychiatry* 11:16. doi: 10.3389/fpsyt.2020.00016

Pagsberg, A. K., Baaré, W. F. C., Raabjerg Christensen, A. M., Fagerlund, B., Hansen, M. B., Labianca, J., et al. (2007). Structural brain abnormalities in early onset first-episode psychosis. *J. Neural Transm.* 114, 489–498. doi: 10.1007/s00702-006-0573-8

Palaniyappan, L., Park, B., Balain, V., Dangi, R., and Liddle, P. (2015). Abnormalities in structural covariance of cortical gyrification in schizophrenia. *Brain Struct. Funct.* 220, 2059–2071. doi: 10.1007/s00429-014-0772-2

Peng, C. J., Lee, K. L., and Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *J. Educ. Res.* 96, 3–14.

Qureshi, M. N. I., Oh, J., and Lee, B. (2019). 3D-CNN based discrimination of schizophrenia using resting-state fMRI. *Artific. Intell. Med.* 98, 10–17. doi: 10.1016/j.artmed.2019.06.003

Rampisela, T. V., and Rustam, Z. (2018). Classification of schizophrenia data using support vector machine (SVM). *J. Phys. Conf. Ser.* 1108:12044.

Ren, W., Lui, S., Deng, W., Li, F., Li, M., Huang, X., et al. (2013). Anatomical and functional brain abnormalities in drug-naive first-episode schizophrenia. *Am. J. Psychiatry* 170, 1308–1316. doi: 10.1176/appi.ajp.2013.12091148

Samartzis, L., Dima, D., Fusar-Poli, P., and Kyriakopoulos, M. (2014). White matter alterations in early stages of schizophrenia: a systematic review of diffusion tensor imaging studies. *J. Neuroimaging* 24, 101–110. doi: 10.1111/j.1552-6569.2012.00779.x

Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* 82, 403–415. doi: 10.1016/j.neuroimage.2013.05.081

Shi, D., Li, Y., Zhang, H., Yao, X., Wang, S., Wang, G., et al. (2021). Machine learning of schizophrenia detection with structural and functional neuroimaging. *Dis. Mark.* 2021, 1–12. doi: 10.1155/2021/9963824

Skåtun, K. C., Kaufmann, T., Doan, N. T., Alnæs, D., Córdova-Palomera, A., Jönsson, E. G., et al. (2017). Consistent functional connectivity alterations in schizophrenia spectrum disorder: a multisite study. *Schizophr. Bull.* 43, 914–924. doi: 10.1093/schbul/sbw145

Steardo, L., Carbone, E. A., de Filippis, R., Pisanu, C., Segura-Garcia, C., Squassina, A., et al. (2020). Application of support vector machine on fMRI data as biomarkers in schizophrenia diagnosis: a systematic review. *Front. Psychiatry* 11:588. doi: 10.3389/fpsyt.2020.00588

Su, L., Wang, L., Shen, H., Feng, G., and Hu, D. (2013). Discriminative analysis of non-linear brain connectivity in schizophrenia: an fMRI Study. *Front. Hum. Neurosci.* 7:702. doi: 10.3389/fnhum.2013.00702

Sui, J., Castro, E., He, H., Bridwell, D., Du, Y., Pearlson, G. D., et al. (2014). Combination of FMRI-SMRI-EEG data improves discrimination of schizophrenia patients by ensemble feature selection. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2014, 3889–3892. doi: 10.1109/EMBC.2014.6944473

Sun, Y., Chen, Y., Collinson, S. L., Bezerianos, A., and Sim, K. (2015). Reduced hemispheric asymmetry of brain anatomical networks is linked to schizophrenia: a connectome study. *Cereb. Cortex* 255, 1–14. doi: 10.1093/cercor/bhv255

Talpalaru, A., Bhagwat, N., Devenyi, G. A., Lepage, M., and Chakravarty, M. M. (2019). Identifying schizophrenia subgroups using clustering and supervised learning. *Schizophr. Res.* 214, 51–59. doi: 10.1016/j.schres.2019.05.044

Torres, U. S., Duran, F. L. S., Schaufelberger, M. S., Crippa, J. A. S., Louzã, M. R., Sallet, P. C., et al. (2016). Patterns of regional gray matter loss at different stages of schizophrenia: a multisite, cross-sectional VBM study in first-episode and chronic illness. *NeuroImage Clin.* 12, 1–15. doi: 10.1016/j.nicl.2016.06.002

Tu, P., Lee, Y., Chen, Y., Li, C., and Su, T. (2013). Schizophrenia and the brain's control network: aberrant within- and between-network connectivity of the frontoparietal network in schizophrenia. *Schizophr. Res.* 147, 339–347. doi: 10.1016/j.schres.2013.04.011

Turner, J. A. (2013). A multi-site resting state fMRI study on the amplitude of low frequency fluctuations in schizophrenia. *Front. Neurosci.* 7:137. doi: 10.3389/fnins.2013.00137

van den Heuvel, M. P., and Fornito, A. (2014). Brain networks in schizophrenia. *Neuropsychol. Rev.* 24, 32–48.

van den Heuvel, M. P., and Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534. doi: 10.1016/j.euroneuro.2010.03.008

van den Heuvel, M. P., Mandl, R. C. W., Stam, C. J., Kahn, R. S., and Hulshoff Pol, H. E. (2010). Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. *J. Neurosci.* 30, 15915–15926. doi: 10.1523/JNEUROSCI.2874-10.2010

van den Heuvel, M. P., Sporns, O., Collin, G., Scheewe, T., Mandl, R. C. W., Cahn, W., et al. (2013). Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry* 70, 783–792. doi: 10.1001/jamapsychiatry.2013.1328

van Tol, M., van der Meer, L., Bruggeman, R., Modinos, G., Knegtering, H., and Aleman, A. (2014). Voxel-based gray and white matter morphometry correlates of hallucinations in schizophrenia: the superior temporal gyrus does not stand alone. *NeuroImage Clin.* 4, 249–257. doi: 10.1016/j.nicl.2013.12.008

Wang, H., Jin, X., Zhang, Y., and Wang, J. (2016). Single-subject morphological brain networks: connectivity mapping, topological characterization and test-retest reliability. *Brain Behav.* 6:e00448. doi: 10.1002/brb3.448

Wang, J., Wang, X., Xia, M., Liao, X., Evans, A., and He, Y. (2015). GRETNA: a graph theoretical network analysis toolbox for imaging connectomics. *Front. Hum. Neurosci.* 9:386. doi: 10.3389/fnhum.2015.00386

Wang, Q., Su, T., Zhou, Y., Chou, K., Chen, I., Jiang, T., et al. (2012). Anatomical insights into disrupted small-world networks in schizophrenia. *NeuroImage* 59, 1085–1093. doi: 10.1016/j.neuroimage.2011.09.035

Wang, S., Zhan, Y., Zhang, Y., Lyu, L., Lyu, H., Wang, G., et al. (2018). Abnormal long- and short-range functional connectivity in adolescent-onset schizophrenia patients: a resting-state fMRI study. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 81, 445–451. doi: 10.1016/j.pnpbp.2017.08.012

Watanabe, T., Kessler, D., Scott, C., Angstadt, M., and Sripada, C. (2014). Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine. *NeuroImage* 96, 183–202. doi: 10.1016/j.neuroimage.2014.03.067

Winterburn, J. L., Voineskos, A. N., Devenyi, G. A., Plitman, E., de la Fuente-Sandoval, C., Bhagwat, N., et al. (2019). Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study. *Schizophr. Res.* 214, 3–10. doi: 10.1016/j.schres.2017.11.038

Wu, F., Zhang, Y., Yang, Y., Lu, X., Fang, Z., Huang, J., et al. (2018). Structural and functional brain abnormalities in drug-naive, first-episode, and chronic patients with schizophrenia: a multimodal MRI study. *Neuropsychiatr. Dis. Treat.* 14, 2889–2904. doi: 10.2147/NDT.S174356

Xiao, Y., Yan, Z., Zhao, Y., Tao, B., Sun, H., Li, F., et al. (2017). Support vector machine-based classification of first episode drug-naïve schizophrenia patients

and healthy controls using structural MRI. *Schizophr. Res.* 214, 11–17. doi: 10.1016/j.schres.2017.11.037

Yamasue, H., Iwanami, A., Hirayasu, Y., Yamada, H., Abe, O., Kuroki, N., et al. (2004). Localized volume reduction in prefrontal, temporolimbic, and paralimbic regions in schizophrenia: an MRI parcellation study. *Psychiatry Res.* 131, 195–207. doi: 10.1016/j.pscychresns.2004.05.004

Yan, C., Wang, X., Zuo, X., and Zang, Y. (2016). DPABI: data processing & analysis for (resting-state) brain imaging. *Neuroinformatics* 14, 339–351.

Zang, J., Huang, Y., Kong, L., Lei, B., Ke, P., Li, H., et al. (2021). Effects of brain atlases and machine learning methods on the discrimination of schizophrenia patients: a multimodal MRI study. *Front. Neurosci.* 15:697168. doi: 10.3389/fnins.2021.697168

Zeng, L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., et al. (2018). Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. *EBioMedicine* 30, 74–85. doi: 10.1016/j.ebiom.2018.03.017

Zhang, Y., Lin, L., Lin, C., Zhou, Y., Chou, K., Lo, C., et al. (2012). Abnormal topological organization of structural brain networks in schizophrenia. *Schizophr. Res.* 141, 109–118. doi: 10.1016/j.schres.2012.08.021

Zhang, Y., Xu, L., Hu, Y., Wu, J., Li, C., Wang, J., et al. (2019). Functional connectivity between sensory-motor subnetworks reflects the duration of untreated psychosis and predicts treatment outcome of first-episode drug-naïve schizophrenia. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 697–705. doi: 10.1016/j.bpsc.2019.04.002

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Ann. Transl. Med.* 4:218. doi: 10.21037/atm.2016.03.37

Zhou, Y., Shu, N., Liu, Y., Song, M., Hao, Y., Liu, H., et al. (2007). Altered resting-state functional connectivity and anatomical connectivity of hippocampus in schizophrenia. *Schizophr. Res.* 100, 120–132. doi: 10.1016/j.schres.2007.11.039

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

frontiers in Neuroscience

OPINION
published: 12 January 2022
doi: 10.3389/fnins.2021.782918

Potential of Overcomplete Wavelet Frame Expansion for Facilitating Electroencephalogram Information Mining

Wanshan Liu[1,2], *Xiaoyue Guo*[1,2], *Binqiang Chen*[1,2*] *and Wangpeng He*[3]

[1] School of Aerospace Engineering, Xiamen University, Xiamen, China, [2] Shenzhen Research Institute of Xiamen University, Shenzhen, China, [3] School of Aerospace Science and Technology, Xidian University, Xi'an, China

Keywords: wavelet transform, EEG, brain imaging, feature extraction, overcomplete wavelet frame expansion

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Linkai Niu,
Taiyuan University of Technology, China
Mingquan Zhang,
Xi'an Jiaotong University, China

****Correspondence:***
Binqiang Chen
cbq@xmu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *25 September 2021*
Accepted: *10 December 2021*
Published: *12 January 2022*

Citation:
Liu W, Guo X, Chen B and He W (2022) Potential of Overcomplete Wavelet Frame Expansion for Facilitating Electroencephalogram Information Mining. Front. Neurosci. 15:782918. doi: 10.3389/fnins.2021.782918

INTRODUCTION

Electroencephalogram (EEG) is a record obtained by amplifying and recording the electrical activity, produced by electrical flows in the brain, from the human scalp (Zandi et al., 2011; Larson and Taulu, 2018). EEG is a widely used medium in brain imaging science and plays an important role in the research of the brain–computer interface (BCI; Gao et al., 2021). BCI is an online computerized system that converts brain signals into useful commands. To date, different types brain signals have been employed to develop BCI systems. Because of its convenience and low cost, EEG signal has become the main medium in BCI systems. However, it has been proved in practice that the acquisition of EEG signal is easily disturbed by various types of noises due to the weak energy of EEG signals. In order to extract useful information from noisy EEG signals (Shad et al., 2020), various signal processing methods are investigated in EEG signal analysis.

In the analysis of brain signals, improving the signal to noise ratio is an important preprocessing step. Traditionally, it is completed using Fast Fourier Transform (FFT) (Wahab et al., 2021). In BCI, FFT is also utilized to implement salient feature extraction from EEG signals. The short-time Fourier transform is an enhancement of FFT and it can generate two-dimensional spectral representation of EEG (Ha and Jeong, 2019). However, the main drawback of STFT is that its frequency resolution is not tunable. Huang proposed a methodology, combining STFT and convolutional neural networks for biomedical signal classification (Huang et al., 2019). In addition, the digital filters based on Fourier analysis is also an important tool for EEG signal denoising (Hsia and Kraft, 1983). Their applications include noisy artifact removal, feature selection at specific frequency bands. Although new techniques for EEG filtering are still emerging recently, the filtering technique is not important focus of BCI research. Shortcomings of digital filters are also reported in related studies (Alhammadi and Mahmoud, 2016).

During the past decades, with the increase of computing power, many more advanced signal processing methods have been invented and put into practice. Upadhyay put forward a novel technique, by integration of S transform and independent component analysis, for artifact removal and noise suppression in EEG signals (Upadhyay et al., 2016). Djemili utilized empirical mode decomposition to decompose EEG signal into intrinsic mode functions and achieved intelligent classification of normal and epileptic EEG features (Djemili et al., 2016). In the study by Jiang, a multi-dictionary based sparse representation approach is proposed for automatic detection of epileptic EEG spikes (Jiang et al., 2020). Dora applied variational mode decomposition for correcting artifacts in EEG measurements (Dora and Biswal, 2020). Chen proposed a sparse Fourier transform and applied it in power-line artifact removal (Chen et al., 2021b).

Although there are a lot of tools for EEG signal analysis, there is still a need to select practical tools to study and improve their performance. We believe that a good EEG signal processing method should have three advantages. Firstly, the method should have a rigorous mathematical basis and can be easily improved theoretically. Secondly, the method has been widely used in clinical practice, and some mature and practical technical solutions have been formed. Thirdly, the method should be computationally efficient and can be deployed quickly with conventional hardware.

JOINT OF WT AND ARTIFICIAL INTELLIGENCE FOR INTELLIGENT ANALYSIS OF EEG

Considering the above three feasible merits of a practical EEG signal analyzing tool, wavelet transform (WT) has become the most commonly used tool for EEG signal analysis. Wavelet transform is a modern development of Fourier analysis. WT can not only extract high-dimensional features from EEG, but also has high computational efficiency (Khatkar and Kumar, 2015). During the past two decades, WT has been successfully applied in EEG feature extraction and noisy reduction. Sartoretto detected features associated with epileptiform activity from EEG signals *via* discrete wavelet analysis (Sartoretto and Ermani, 1999). Mamun explored the utilization of wavelet denoising in physiological noise removal of EEG (Mamun et al., 2013). Ma proposed a method for coherence analysis, between EEG and EMG, based on wavelet decomposition (Ma et al., 2014). Asadpour designed a 4-layer Symmlet-8 wavelet transform structure for EEG signals, and effectively decomposed δ, θ, α, and β brain rhythm waves into different subspaces (Asadpour et al., 2018). Li employed wavelet packet transform (WPT) to decompose the non-stationary measurement signal in time and frequency domain, and selects the frequency band information related to the imagination task to reconstruct the EEG signal features (Li et al., 2021). Obukhov utilized ridges of wavelet spectra for automatic diagnose of epileptic seizure (Obukhov et al., 2021). It can be seen that the classical wavelet transform is mainly used to decompose EEG signals in the literature, and there is no in-depth study on the impact of wavelet transform on the decomposition results.

The wide application of wavelet transform in EEG signal analysis is not only reflected in the signal decomposition, but also in the information extraction of the decomposed subspaces. A most immediate way is to carry out statistical analysis on the decomposed wavelet subspace to obtain the corresponding statistical feature space. For example, Liu utilized many statistical indicators from selected wavelet subspaces to implement automatic seizure detection (Liu et al., 2012). In recent years, with the deepening of research, many features with physical significance have been designed and used in EEG signal classification. Zhang used sliding window technology to extract wavelet entropy, sample entropy and peak-peak value, and effectively identifies four States of driver fatigue: normal state, mild fatigue, emotional fluctuation and excessive fatigue (Zhang et al., 2014). Hadjileontiadis studied higher order spectral features in the wavelet subspaces, and proposed a novel methodology for characterization of tonic cold pain (Hadjileontiadis, 2015). Peng explored the indicator of wavelet entropy of EEG in fatigue detection, and found it provides better performance compared with FFT based indicators (Peng et al., 2021). Zarei explored nonlinear features in wavelet subspaces to improve the accuracies of automatic seizure detection (Zarei and Asl, 2021).

Although wavelet transform can reveal the time-frequency characteristics of EEG, it will take a lot of time to identify and classify them manually. This is becoming increasingly impossible in the era of medical big data. Therefore, as an important tool for feature extraction, wavelet transform also needs to be combined with artificial intelligence to achieve intelligent analysis results (Cao et al., 2019). Sharma combined wavelet subspace features and support vector machine for EEG driven epilepsy diagnosis (Sharma et al., 2020). Albaqami studied the automatic EEG signal classification using WPT and gradient boosting decision tree (Albaqami et al., 2021). Movahed employed a special orthogonal wavelet filter-bank for EEG decomposition and combined it with machine learning for diagnose the disease of major depressive disorder (Movahed et al., 2021). Shahabi studied drug responses of major depressive disorder using a technique that applied deep transfer learning on wavelet-based features from EEG (Shahabi et al., 2021). In the literature, a large number of research results show that wavelet transform is an indispensable pre-processing tool for artificial intelligence recognition of EEG signals.

STATE-OF-THE-ART DEVELOPMENT OF OVERCOMPLETE WAVELET FRAME EXPANSIONS AND ITS POTENTIAL APPLICATIONS IN EEG ANALYSIS

The above materials show that the discrete wavelet transform is more commonly used in EEG analysis in clinical practice because of its computational efficiency. The use of wavelet transform has also been shown to be beneficial because its association with artificial intelligence can significantly improve the accuracy of clinical distortion diagnosis. Among the various types of discrete wavelet transform, classical wavelet transform and wavelet packet transform are most commonly used. They enjoy the highest computational efficiency because they are conventional linear expansions. However, both of the two signal decomposition tools have some significant shortcomings in theory. These shortcomings include translation sensitivity in wavelet subspaces (Bayram and Selesnck, 2008), fixed frequency-scale pavement (Chen et al., 2021a), and difficulties of wavelet basis information fusion.

To a large extent, the reason for these shortcomings is that the classical wavelet transform is a basis transform. For basis transformation, the number of linearly independent vectors in the basis is the same as the length of the input signal (**Figure 1A**).

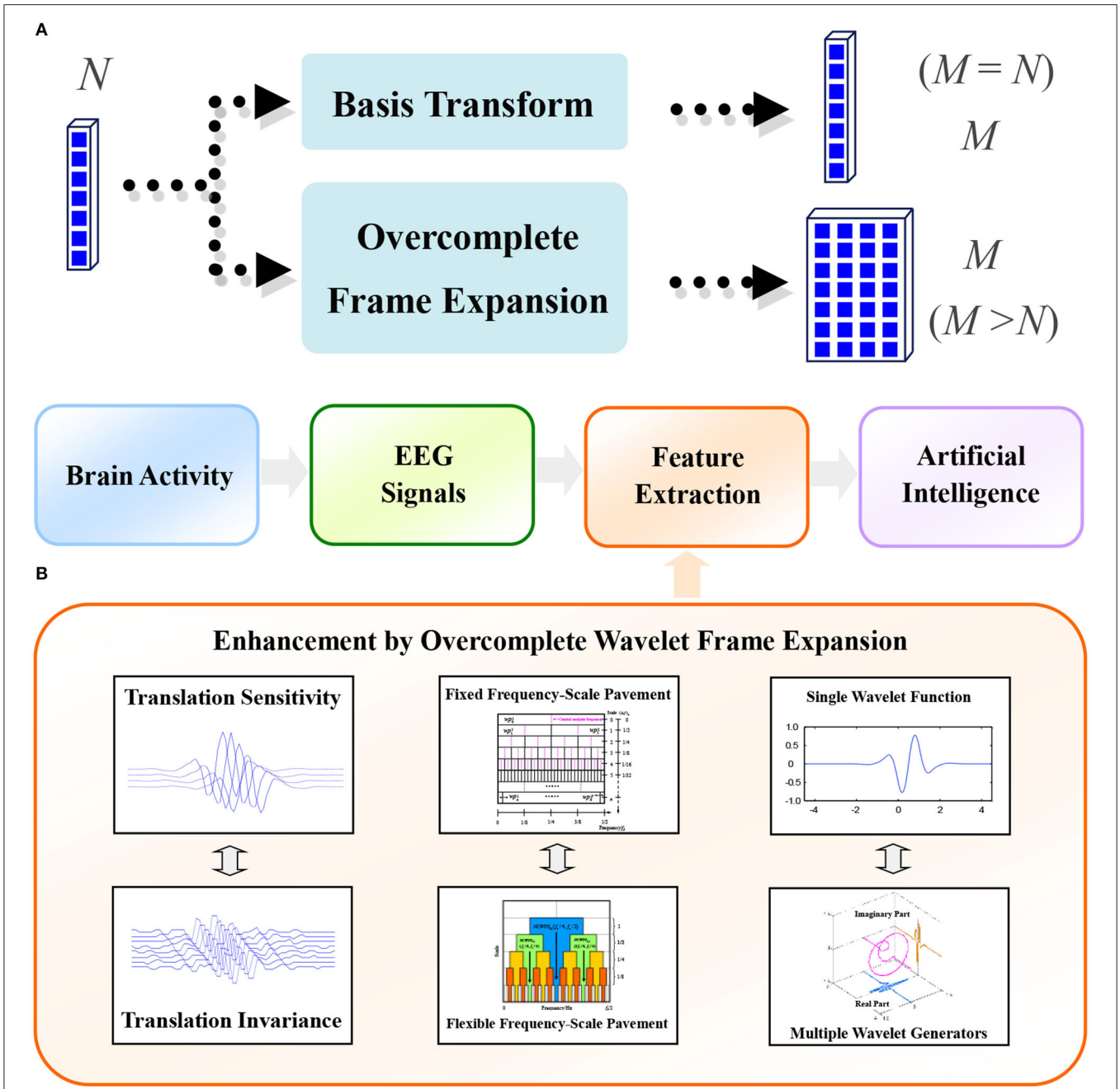


FIGURE 1 | (A) Fundamentals of overcomplete wavelet frame expansion, where N is the number of samples in the input signal and M is the number of samples in the output signal; and **(B)** A framework to illustrate the enhancement of wavelet decomposition for EEG-based information mining.

While, for overcomplete expansion, a set of linearly dependent vectors, whose number is greater than the length of the input signal, is utilized to represent the input signal (**Figure 1A**). This set of linearly dependent vectors is often called a dictionary. By introducing the development of frame theory, these problems can be properly resolved (Kovacevic and Chebira, 2007a,b). A framework for achieving these demands can be found in **Figure 1B**. Scholars have tried to improve the properties of wavelet transform from three aspects. One is to improve the translation invariance of signal expansion, the other is to adjust the "frequency-scale" pavements, and the third is to try to introduce multiple wavelet functions (Chen, 2014).

Translation sensitivity is caused by down-sampling operations in implementation of wavelet filter-bank. Although the property

of perfect reconstruction can be guaranteed for the input original signal, the side effect of feature distortion in wavelet subspaces may be very serious. To address this problem, one direct and simple method is to eliminate the down-sampling operator on the filter bank, but this will significantly increase the computational complexity of the discrete wavelet analysis (Li et al., 2012). Due to the addition of redundancy, other construction constraints can be added to the construction of the wavelet bases. The constraint of Hilbert transform pair can generate approximate translation invariance with relatively small increase in calculation (Huang et al., 2021).

The property of fixed frequency-scale pavement leads to inflexible analysis results, especially when the EEG features are located in the transition region of the classical discrete wavelet passing band. Therefore, inappropriate use of classical WT and WPT will cause serious distortion of the features in the EEG decomposition results. There are two available solutions to the address this problem in the theory of overcomplete frame expansion. One is to add more wavelet functions to the wavelet base to change the basic time-frequency characteristics of the wavelet base (Selesnick, 2004). In this case, the dilation factor of the wavelet transform and the basic structure of the filter bank remain unchanged. Its disadvantage is that the effect of adjusting the frequency-scale pavement is limited, and sometimes it cannot meet the analysis requirements. The other method is to construct the wavelet time-frequency atom completely in the frequency domain by using the analytical expression (Gilles, 2013). In this case, a highly flexible dilation factor can be chosen. Its disadvantage is that there are a large number of parameters to be determined in the construction process. In recent years, some researchers have tried to use the wavelet transform with adjustable dilation factor to decompose EEG signals.

The most difficult problem is the information fusion of multiple wavelet bases with different time-frequency characteristics (He et al., 2017). With the development of sparse representation theory, wavelet bases can be used as the dictionary of signal sparse decomposition, so the information fusion becomes possible. However, the relevant decomposition models are still very limited, so they cannot be well-applied to EEG signal analysis. A more promising solution is to combine the multiple wavelet dictionaries with composite deep learning networks to achieve deep fusion of information through neural networks.

SUMMARY

The classical discrete wavelet transform is widely used in EEG signal analysis. The combination of wavelet transform and artificial intelligence can significantly improve the accuracy of clinical disease diagnosis. However, the influence of wavelet transform on the feature space of EEG signal decomposition is often neglected. At present, with the advancement of wavelet analysis theory, the feature space can be enhanced by using the overcomplete wavelet frame decomposition with better performance. The construction of wavelet bases and the redundancy of filter-banks can be combined with other useful construction constraints to improve the shortcomings of classical discrete wavelet transform. At present, there have been effective solutions to the property of translation sensitivity and the fixed "frequency-scale" pavement. The combination of overcomplete wavelet frame expansion and composite deep learning network can significantly promote the deep fusion of information. We recommend the use of overcomplete wavelet frame expansion in the feature extraction and signal classification of EEG, and suggest that further researches should be carried out. It should be pointed out that although overcomplete frame expansion can provide more flexible time-frequency analysis expansions, it also requires higher computational requirements. Its computation time is usually several times to tens of times that of the classical wavelet transform. Due to the increase of computing resources and the enhancement of computing power in recent years, such demands can be well-satisfied.

AUTHOR CONTRIBUTIONS

BC conceptualized the present work. WL, BC, and XG wrote the manuscript. WH reviewed and edited the manuscript. All authors read and approved the manuscript.

FUNDING

This research was supported financially by National Natural Science Foundation of China (Grant No. 51805398), Project of Youth Talent Lift Program of Shaanxi University Association for Science and Technology (Grant No. 20200408), Fundamental Research Funds for the Central Universities (Grant No. JB211303), and Fundamental Research Funds for the Central Universities under Grant (No. 20720190009).

REFERENCES

Albaqami, H., Hassan, G. M., Subasi, A., and Datta, A. (2021). Automatic detection of abnormal EEG signals using wavelet feature extraction and gradient boosting decision tree. *Biomed. Sign. Proces. Control* 70:102957. doi: 10.1016/j.bspc.2021.102957

Alhammadi, A. A., and Mahmoud, S. A. (2016). Fully differential fifth-order dual-notch powerline interference filter oriented to EEG detection system with low pass feature. *Microelectr. J.* 56, 122–133. doi: 10.1016/j.mejo.2016.08.014

Asadpour, A., Jahed, M., and Mahmoudian, S. (2018). Brain waves evaluation of sound therapy in chronic subjective tinnitus cases using wavelet decomposition. *Front. Integr. Neurosci.* 12:38. doi: 10.3389/fnint.2018.00038

Bayram, I., and Selesnck, I. W. (2008). On the dual-tree complex wavelet packet and M-band transforms. *IEEE Trans. Sign. Proces.* 56, 2298–2310. doi: 10.1109/TSP.2007.916129

Cao, X. C., Chen, B. Q., Yao, B., and He, W. P. (2019). Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification. *Comput. Indus.* 106, 71–84. doi: 10.1016/j.compind.2018.12.018

Chen, B. Q., Li, Y., Cao, X. C., Sun, W., and Zhuang, S. (2021a). Recovery of undersampled force measurement from high-speed milling process using approximate sparsity in frequency domain. *Measurement* 175:109143. doi: 10.1016/j.measurement.2021.109143

Chen, B. Q., Zheng, B. X., Wang, C. Q., and Sun, W. F. (2021b). Adaptive sparse detector for suppressing powerline component in EEG measurements. *Front. Public Health* 9:669190. doi: 10.3389/fpubh.2021.669190

Chen, G. Y. (2014). Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features. *Expert Syst. Appl.* 41, 2391–2394. doi: 10.1016/j.eswa.2013.09.037

Djemili, R., Bouruoba, H., and Korba, M. C. A. (2016). Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals. *Biocybernet. Biomed. Eng.* 36, 285–291. doi: 10.1016/j.bbe.2015.10.006

Dora, C., and Biswal, P. K. (2020). Correlation-based ECG artifact correction from single channel EEG using modified variational mode decomposition. *Comput. Method. Progr. Biomed.* 183:105092. doi: 10.1016/j.cmpb.2019.105092

Gao, X. R., Wang, Y. J., Chen, X. G., and Gao, S. (2021). Interface, interaction, and intelligence in generalized brain–computer interfaces. *Trends Cogn. Sci.* 25, 674–684. doi: 10.1016/j.tics.2021.04.003

Gilles, J. (2013). Empirical wavelet transform. *IEEE Trans. Sign. Proces.* 61, 3999–4010. doi: 10.1109/TSP.2013.2265222

Ha, K. W., and Jeong, J. W. (2019). Motor imagery EEG classification using capsule networks. *Sensors* 19:2854. doi: 10.3390/s19132854

Hadjileontiadis, L. J. (2015). EEG-based tonic cold pain characterization using wavelet higher order spectral features. *IEEE Trans. Biomed. Eng.* 62, 1981–1991. doi: 10.1109/TBME.2015.2409133

He, W. P., Yang, L., Wang, F., and Huang, S. P. (2017). Identification of epileptic EEG signals based on the tunable Q-factor wavelet transform. *J. Biomed. Eng. Res.* 36, 346–350. doi: 10.19529/j.cnki.1672-6278.2017.04.14

Hsia, T. C., and Kraft, R. H. (1983). A digital signal processing system for EEG frequency analysis. *Comput. Progr. Biomed.* 17, 27–33. doi: 10.1016/0010-468X(83)90023-5

Huang, J. S., Chen, B. Q., Li, Y., and Sun, W. F. (2021). Fractal geometry of wavelet decomposition in mechanical signature analysis. *Measurement* 173:108571. doi: 10.1016/j.measurement.2020.108571

Huang, J. S., Chen, B. Q., Yao, B., and He, W. P. (2019). ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access* 7, 92871–92880. doi: 10.1109/ACCESS.2019.2928017

Jiang, Y., Chen, W. Z., Zhang, T., Li, M., You, Y., Zheng, X., et al. (2020). Developing multi-component dictionary-based sparse representation for automatic detection of epileptic EEG spikes. *Biomed. Sign. Proces. Control* 60: 101966. doi: 10.1016/j.bspc.2020.101966

Khatkar, K., and Kumar, D. (2015). Biomedical image enhancement using wavelets. *Proc. Comput. Sci.* 48, 513–517. doi: 10.1016/j.procs.2015.04.128

Kovacevic, J., and Chebira, A. (2007a). Life beyond bases: the advent of frames (part I). *IEEE Sign. Proces. Magazine* 24, 86–104. doi: 10.1109/MSP.2007.4286567

Kovacevic, J., and Chebira, A. (2007b). Life beyond bases: the advent of frames (Part II). *IEEE Sign. Proces. Magazine* 24, 115–125. doi: 10.1109/MSP.2007.904809

Larson, E., and Taulu, S. (2018). Reducing sensor noise in MEG and EEG recordings using oversampled temporal projection. *IEEE Trans. Biomed. Eng.* 65, 1002–1013. doi: 10.1109/TBME.2017.2734641

Li, N., Zhou, R., Hu, Q. H., and Liu, X. H. (2012). Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine. *Mech. Syst. Sign. Proces.* 2, 608–621. doi: 10.1016/j.ymssp.2011.10.016

Li, R. L., Cheng, L. W., Yu, G. J., Zhang, Z. H., and Yu, S. Y. (2021). An electroencephalogram signal processing method fusing wavelet packet and neural network. *J. Beijing Univ. Posts Telecommun.* 3, 94–99. doi: 10.13190/j.jbupt.2020-208

Liu, Y. X., Zhou, W. D., Yuan, Q., and Chen, S. (2012). Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 749–755. doi: 10.1109/TNSRE.2012.2206054

Ma, P. P., Chen, Y. Y., Du, Y. H., Su, Y., Wu, X., Liang, Z., et al. (2014). Analysis of corticomuscular coherence during rehabilitation exercises after stroke. *J. Biomed. Eng.* 5, 971–977. doi: 10.7507/1001-5515.20140183

Mamun, M., Al-Kadi, M., and Marufuzzaman, M. (2013). Effectiveness of wavelet denoising on electroencephalogram signals. *J. Appl. Res. Technol.* 11, 156–160. doi: 10.1016/S1665-6423(13)71524-4

Movahed, R. A., Jahromi, G. P., Shahyad, S., and Meftahi, G. H. (2021). A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis. *J. Neurosci. Method.* 358:109209. doi: 10.1016/j.jneumeth.2021.109209

Obukhov, Y. V., Kershner, I. A., Tolmacheva, R. A., Sinkin, M. V., and Zhavoronkova, L. A. (2021). Wavelet ridges in EEG diagnostic features extraction: epilepsy long-time monitoring and rehabilitation after traumatic brain injury. *Sensors* 21:5989. doi: 10.3390/s21185989

Peng, Y. F., Wong, C. M., Wang, Z., Rosa, A. C., Wang, H. T., Wan, F., et al. (2021). Fatigue detection in SSVEP-BCIs based on wavelet entropy of EEG. *IEEE Access* 9, 114905–114913. doi: 10.1109/ACCESS.2021.3100478

Sartoretto, F., and Ermani, M. (1999). Automatic detection of epileptiform activity by single-level wavelet analysis. *Clin. Neurophysiol.* 110, 239–249. doi: 10.1016/S0013-4694(98)00116-3

Selesnick, I. W. (2004). The double-density dual-tree DWT. *IEEE Trans. Sign. Proces.* 52, 1304–1314. doi: 10.1109/TSP.2004.826174

Shad, E. H. T., Molinas, M., and Ytterdal, T. (2020). Impedance and noise of passive and active dry EEG electrodes: a review. *IEEE Sens. J.* 20, 14565–14577. doi: 10.1109/JSEN.2020.3012394

Shahabi, M. S., Shalbaf, A., and Maghsoudi, A. (2021). Prediction of drug response in major depressive disorder using ensemble of transfer learning with convolutional neural network based on EEG. *Biocyberneti. Biomed. Eng.* 41, 946–959. doi: 10.1016/j.bbe.2021.06.006

Sharma, M., Patel, S., and Acharya, U. R. (2020). Automated detection of abnormal EEG signals using localized wavelet filter banks. *Pattern Recognit. Lett.* 133, 188–194. doi: 10.1016/j.patrec.2020.03.009

Upadhyay, R., Padhy, P. K., and Kankar, P. K. (2016). EEG artifact removal and noise suppression by discrete orthonormal S-transform denoising. *Comput. Electr. Eng.* 53, 125–142. doi: 10.1016/j.compeleceng.2016.05.015

Wahab, M. F., Gritti, F., and O'Haver, T. C. (2021). Discrete Fourier transform techniques for noise reduction and digital enhancement of analytical signals. *Trends Analyt. Chem.* 143:116354. doi: 10.1016/j.trac.2021.116354

Zandi, A. S., Dumont, G. A., Yedlin, M. J., Lapeyrie, P., Sudre, C., and Gaffet, S. (2011). Scalp EEG acquisition in a low-noise environment: a quantitative assessment. *IEEE Trans. Biomed. Eng.* 58, 2407–2417. doi: 10.1109/TBME.2011.2158647

Zarei, A., and Asl, B. M. (2021). Automatic seizure detection using orthogonal matching pursuit, discrete wavelet transform, and entropy based features of EEG signals. *Comput. Biol. Med.* 131:104250. doi: 10.1016/j.compbiomed.2021.104250

Zhang, C., Wang, H., and Fu, R. (2014). Automated detection of driver fatigue based on entropy and complexity measures. *IEEE Trans. Intelli. Transport. Syst.* 15, 168–177. doi: 10.1109/TITS.2013.2275192

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

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

Copyright © 2022 Liu, Guo, Chen and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 14 January 2022
doi: 10.3389/fnins.2021.829040

Sparse Representation-Based Discriminative Metric Learning for Brain MRI Image Retrieval

*Guohua Zhou[1,2,3], Bing Lu[1], Xuelong Hu[3] and Tongguang Ni[2]**

[1] School of Information Engineering, Changzhou Institute of Industry Technology, Changzhou, China, [2] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China, [3] College of Information Engineering, Yangzhou University, Yangzhou, China

Magnetic resonance imaging (MRI) can have a good diagnostic function for important organs and parts of the body. MRI technology has become a common and important disease detection technology. At the same time, medical imaging data is increasing at an explosive rate. Retrieving similar medical images from a huge database is of great significance to doctors' auxiliary diagnosis and treatment. In this paper, combining the advantages of sparse representation and metric learning, a sparse representation-based discriminative metric learning (SRDML) approach is proposed for medical image retrieval of brain MRI. The SRDML approach uses a sparse representation framework to learn robust feature representation of brain MRI, and uses metric learning to project new features into the metric space with matching discrimination. In such a metric space, the optimal similarity measure is obtained by using the local constraints of atoms and the pairwise constraints of coding coefficients, so that the distance between similar images is less than the given threshold, and the distance between dissimilar images is greater than another given threshold. The experiments are designed and tested on the brain MRI dataset created by Chang. Experimental results show that the SRDML approach can obtain satisfactory retrieval performance and achieve accurate brain MRI image retrieval.

Keywords: medical image retrieval, magnetic resonance imaging, brain images, sparse representation, metric learning

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Ming Gao,
Wuhan Sports University, China
Tao Tao,
Anhui University of Technology, China

****Correspondence:***
Tongguang Ni
ntg@cczu.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *04 December 2021*
Accepted: *27 December 2021*
Published: *14 January 2022*

Citation:
Zhou G, Lu B, Hu X and Ni T (2022) Sparse Representation-Based Discriminative Metric Learning for Brain MRI Image Retrieval. Front. Neurosci. 15:829040. doi: 10.3389/fnins.2021.829040

INTRODUCTION

The number of tumor patients has increased rapidly in recent years. Tumors have become one of the most common diseases in the world. The prevention, early diagnosis, and treatment of tumors have aroused widespread concern and discussion in the medical community and the public. Pathological examination is still the gold standard for tumor detection in the medical field. Medical imaging is a common method for diagnosing tumors, such as X-ray, magnetic resonance imaging (MRI), and computed tomography (CT) scans (Zhang et al., 2019; Chahal et al., 2020; Jiang et al., 2021). At present, early cancer screening mainly depends on the experience of doctors. The accuracy of decision-making largely depends on the knowledge and experience of doctors and the quality and

quantity of information available. At the same time, the development of digital imaging technology and artificial intelligence has jointly promoted the emergence of digital pathology. Digital pathology aims to use computers to automatically analyze the characteristics of digital histopathology images to serve different tasks such as detection, segmentation, and retrieval. Digital pathology can provide physicians with knowledge-supported assistive technology in their decision-making process, which is of great practical significance.

With the popularization of Internet technology, the scale of medical imaging databases is getting larger and larger, and more and more past cases can be provided to doctors. Content-based image retrieval (CBIR) has technical advantages in medical diagnosis (Müller and Unay, 2017; Zheng et al., 2018). CBIR not only provides doctors with a prediction, but also improves the efficiency and accuracy of diagnosis by retrieving similar historical cases. Therefore, CBIR is essentially consistent with the physician's diagnostic decision-making process. In addition, the use of CBIR can develop knowledge-based medical diagnosis decision support systems and case interpretation systems. Medical image retrieval is the core technology of the CBIR system, which can retrieve medical images similar to the image to be diagnosed from the medical database. We know that medical images are very different from natural images. First, the resolution of medical images is relatively high, but most of them are grayscale images. Secondly, the important information of medical images is mostly concentrated in small areas. Thirdly, the semantic content of visually similar medical images may be very different. Therefore, the retrieval effect is often unsatisfactory when the conventional natural image processing method is directly applied to medical imaging. Especially for brain tumor images, the shape, size, and texture of the same type of brain tumor are different due to the severity of the patient's condition, age, and other factors. Therefore, different pathological types may show similar appearances, which make the recognition of brain tumor images very challenging.

Medical image retrieval is to retrieve images similar to the histopathological images to be diagnosed from the database, to provide references for doctors' diagnoses. The histopathological image retrieval algorithm generally includes two stages: feature extraction and similarity measurement model construction (Zhang et al., 2015). The tumor boundary in brain tumor images is usually unclear, so the shape information cannot provide a clear tumor contour area. Therefore, it is important to determine the intensity and texture features. Huang et al. (2012) combined the spatial pyramid model and the Bow model and used the bag-of-words (BoW) histogram with spatial information to describe the characteristics of brain tumor regions, which improved the retrieval performance of brain tumor images. Huang et al. (2014) regarded the boundary of the brain tumor as the region of interest (ROI), used the region division learning method, and expressed the features with the original image pixel intensity. Finally, the authors aggregated the local features of each subregion and spatialized them, to improve the discriminative ability of image features. Cheng et al. (2016) adopted adaptive space pool and fisher vector method for brain tumor image retrieval.

In the field of medical image retrieval, similarity measurement refers to using a given medical image pair to calculate the distance between the feature vectors of medical image pairs, and it is often used to judge the type of medical image. For medical images, similarity measures include semantic correlation and visual similarity. The semantic correlation measure depends on the class labels of medical images. For example, the labels of two brain MRI images are healthy, which means that the two brain MRI images are semantically related. The visual similarity measure describes the similarity of features, which usually describes the similarity of medical images from a visual perspective. The definition of similarity measures generally considers the use of distance measures. The commonly used distance measurement methods include Euclidean distance, angle cosine distance, Mahalanobis distance, and Minkowski distance, etc., Yang et al. (2012) extracted local features based on pixel intensity information along the brain tumor boundary and used BoW to generate global features. At the same time, the authors designed a distance metric learning method to improve retrieval accuracy. Verma and Jawahar (2017) proposed a two-pass k-nearest neighbor model. Firstly, a more balanced training data set was designed for each image to reduce the label imbalance problem. Then, the large interval nearest neighbor algorithm was extended to the case of multi-label classification for metric learning, and the optimal weights of combined basis distance and features were obtained for semantic label prediction. Gu et al. (2017) proposed a distance measurement method of weighted heterogeneous values for tumor diagnosis.

Inspired by the sparse mechanism of the human visual system, sparse representation has been widely used in the field of computer vision. Sparse representation succinctly represents the image to be labeled as a linear combination of a few atoms in the dictionary. These sparse representation coefficients and corresponding atoms fully reveal the internal essence of the image and also adapted to human visual perception. Dictionary learning is a common method in sparse representation research. The dictionary learned from a large number of samples can adaptively sparse represent various feature information in the image, which greatly improves the sparse representation performance. Sparse representation methods can be divided into two categories. The first is methods based on predefined dictionaries. Srinivas et al. (2014) used the correlation between the RGB three-channels of color images to construct a dictionary and applied it to histopathological image representation and classification. However, this method based on the predefined dictionary may not be able to make full use of the discriminative information hidden in the training samples. The second type of method learns the dictionaries from training samples. For example, Gao et al. (2014) regarded the samples of each subclass as a subgroup and used the multi-layer group sparsity to classify and annotate images. Tang et al. proposed a semi-supervised learning method based on a sparse graph for image classification. By using the multi-scale and spatial information of the image, Tang et al. (2010) proposed a multi-scale representation learning algorithm for breast image classification. Zhang et al. (2015) proposed a heterogeneous feature fusion method and applied it to the image classification of pathological tissues.

The above methods mainly focus on the reconstruction of histopathological images, so their classification performances are not optimal. The dictionary learning method based on supervised learning uses robust feature representation from the original visual information through mining supervised or semi-supervised features and coding coefficients. Cao et al. (2015) used the group sparse reconstruction method to learn the semantic correlation of symbiotic labels for optimizing the dictionary. The authors used the group sparse framework to reconstruct the test image for label prediction and realized weakly supervised image annotation. Lu et al. (2015) studied the semantic sparse recording method of visual content to generate a more descriptive and robust bow representation for image classification.

This paper proposes sparse representation-based discriminative metric learning (SRDML) approach for brain MRI image retrieval. SRDML integrates sparse representation and metric learning into a discriminative model. On the one hand, the model uses the sparse representation to learn the robust feature representation of brain MRI images. On the other hand, the model learns a metric space with discriminative ability, so that similar brain MRI images are closely projected in the metric space, and dissimilar brain MRI images are separated from each other as much as possible. The advantages of SRDML are as follows: (1) the local information retention term of coding coefficients maintains the semantic correlation and visual similarity of brain MRI images in the projection space. (2) Our approach finds an appropriate metric matrix under the constraints of similarity, so that the distance between similar images' coding coefficients is less than the given threshold, and the distance between dissimilar images' coding coefficients is greater than the given threshold. Its goal is to achieve the dispersion of "maximum inter-class" and "minimum intra-class". (3) The learning of sparse representation and metric learning is optimized through alternate iterations. This learning strategy allows the model parameters to reach the optimal solution at the same time. (4) The experimental results show that SRDML has achieved good performance on the brain MRI image dataset, which shows the feasibility of our retrieval approach in the brain tumors diagnosis.

The rest of this paper is organized as follows: first, related works of metric learning and sparse representation are introduced. Second, the proposed sparse representation-based discriminative metric learning approach is formulated in section "Sparse Representation-Based Discriminative Metric Learning". Third, experiments are presented to verify our approach. Finally, a summary of our approach is presented, and its future work is discussed.

BACKGROUNDS

Metric Learning

The traditional metric learning is based on labeled samples of medical image data sets and the label information appears in the form of paired constraints of samples. Assuming that the medical image data set is expressed as $\{x_i\}_{i=1}^{N} \in R^d$, where x_i is the ith sample, taking Mahalanobis distance as an example, the distance measurement between two images x_i and x_j can be written as,

$$\mathrm{d}_{\mathrm{M}}^2(\mathrm{x_i}, \mathrm{x_j}) = \|\mathrm{x_i} - \mathrm{x_j}\|_{\mathrm{M}}^2 = (\mathrm{x_i} - \mathrm{x_j})^{\mathrm{T}}\mathrm{M}(\mathrm{x_i} - \mathrm{x_j}). \quad (1)$$

The positive semi-definite matrix **M** can be decomposed into $M = W^T W$, and the matrix $W \in R^{d \times m}(m \leq d)$is called the metric matrix. Therefore, Eq. 1 can be expressed as,

$$\mathrm{d}_{\mathrm{M}}^2(\mathrm{x_i}, \mathrm{x_j}) = (\mathrm{Wx_i} - \mathrm{Wx_j})^{\mathrm{T}}(\mathrm{Wx_i} - \mathrm{Wx_j}). \quad (2)$$

Therefore, the essence of the Mahalanobis metric is to learn a projection space, in which that similar image pairs output a positive value close to zero, and dissimilar image pairs output a larger value.

A type of famous metric learning algorithm is large-margin distance metric learning (Weinberger and Saul, 2009; Nguyen et al., 2020). Let **S** and **D** be two datasets of pairwise constraints: $\mathbf{S} = \{(x_i, x_j) \mid x_i \text{ and } x_j \text{ are similar}\}$, and $\mathbf{D} = \{(x_i, x_j) \mid x_i \text{ and } x_j \text{ are dissimilar}\}$. The large-margin distance metric learning takes the paired similar or dissimilar data sets as model input and constructs the metric learning as a convex optimization problem,

$$\begin{aligned} \min_{M} \quad & \textstyle\sum_{(x_i - x_j) \in S} \|x_i - x_j\|_M^2, \\ s.t. \quad & \textstyle\sum_{(x_i - x_j) \in D} \|x_i - x_j\|_M^2 \geq c, \\ & M \geq 0, \end{aligned} \quad (3)$$

where c is a positive. It can be seen from Eq. 3 that the distance between dissimilar pair samples is greater than the constant c, while the distance between similar pair samples is as small as possible.

Sparse Representation

The goal of sparse representation is to use the linear combination of dictionary atoms to represent the observation data, and to use sparsity constraints on the combination coefficients, so that each observation data is only represented by a subset of all dictionary atoms (Gu et al., 2021a,b). The traditional sparse representation framework is,

$$\begin{aligned} \min_{B,Y} \quad & \|\mathrm{X\text{-}BY}\|_F^2 + \lambda\phi(Y), \\ s.t. \quad & \|b_i\|_2 \leq 1, \ \forall i, \end{aligned} \quad (4)$$

where **B** and **Y** are the dictionary matrix and sparse coding coefficient matrix, respectively. $\|\|_F^2$ is F-norm.

In order to avoid that the value of **B** is too large and the sparse coefficient **Y** is too small, the dictionary column vector is usually set to satisfy the constraint $\|b_i\|_2 \leq 1$. $\phi(Y)$ is the regularization term controlling the sparsity of **Y**. λ is the regularization parameter. The objective function is not a joint convex function of (**B**, **Y**), but when one variable is fixed, the objective function is convex for the other variable. Therefore, Eq. 4 can be solved by a turn-by-turn optimization method.

SPARSE REPRESENTATION-BASED DISCRIMINATIVE METRIC LEARNING

In order to better explain the proposed model, first, we introduce some important terms: (1) a labeled training image set $X = [x_1, x_2, ..., x_n] \in R^{d\times n}$; (2) the testing image set contains images available for the retrieval of the target brain MRI images; (3) each image in testing image set is considered as the query image.

The Pairwise Constraint of Sparse Coding Coefficients

From the perspective of image retrieval, similar images should have similar coding coefficients, and dissimilar images should have different coding coefficients. In other words, in the learned metric space, the smaller the better the distance between similar images, at the same time, the greater the better the distance between dissimilar images, to reduce the uncertainty of misjudging the coding coefficient of dissimilar images as similar images. Based on this idea, two thresholds σ_1 and σ_2 are set for calculating the metric distances of similar images and dissimilar images in the metric space as,

$$\begin{cases} d_M^2(y_i - y_j) < \sigma_1, \ if \ \{x_i, x_j\} \in S \\ d_M^2(y_i - y_j) > \sigma_2, if \ \{x_i, x_j\} \in D \end{cases} \tag{5}$$

where $\sigma_1 < \sigma_2$.

To unify the Eq. 5 into an inequality, a common threshold is introduced, i.e., $\sigma_1 = \tau - 1$ and $\sigma_2 = \tau + 1$, then Eq. 5 can be written as,

$$\ell_{ij}(\tau - d_M^2(y_i - y_j)) > 1, \tag{6}$$

where if x_i and x_j are similar images, $\ell_{ij} = 1$; otherwise, x_i and x_j are dissimilar images, $\ell_{ij} = -1$.

With the pairwise constraint of coding coefficients, it is expected to have a large margin between each similar and dissimilar pairs in the learned metric space. By introducing the generalized logistic loss function $g(y) = \frac{1}{\theta}\log(1 + \exp(\theta y))$ with the sharpness parameter θ, the pairwise constraint of coding coefficients can be defined as,

$$\min_Y \ g\left(1 - \ell_{ij}(\tau - d_M^2(y_i - y_j))\right). \tag{7}$$

The Locality Constraint of Atoms

First, to exploit the local structure information in the sparse representation learning, the local manifold of the dictionary atoms is built on the nearest neighbor graph **P** of dictionary $\mathbf{B} \in R^{d\times K}$. The element P_{ij} of **P** is defined as,

$$P_{ij} = \begin{cases} \exp\left(-\frac{\|b_i - b_j\|_2^2}{\mu}\right), & if \ b_j \in \mathrm{KNN}(b_i) \\ 0, & else \end{cases} \tag{8}$$

where KNN( ) is the *k*-nearest neighbor function, and μ is the KNN parameter.

In the learned metric space, we assume that the similar images will keep close to each other, thus it is expected the similar dictionary atoms will keep close to each other as well. A Laplacian matrix **L** is constructed based on the matrix **P**,

$$L = W - P, \tag{9}$$

where $W = diag(W_1, ..., W_K)$, and $W_i = \sum_j^K P_{ij}$. K is the number of atoms in **B**.

Using the Laplacian matrix **L**, the locality constraint of atoms is represented as,

$$\min_Y \ \frac{1}{2}\sum_i^K \sum_j^K (y_i - y_j)^2 P_{ij} = Tr(Y^T L Y). \tag{10}$$

The Objective Function of SRDML and Its Optimization

Now we have the two terms needed to be embedded into the sparse representation framework and compose the SRDML objective function. SRDML combines the sparse representation, pairwise constraint of coding coefficients, and locality constraint of atoms together to yield an image retrieval approach. The SRDML approach takes advantage of pairwise metric constraints and local structure knowledge preserved provided by the coding coefficients. The objective function of SRDML is,

$$\begin{aligned} \min_{M,B,Y} \ & \|\text{X-BY}\|_2^2 + \alpha Tr(YLY) + \beta \textstyle\sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right) \\ & + \gamma \|Y\|_2^2, \\ s.t. \quad & M \geq 0, \\ & \|b_i\|^2 = 1, \ \forall i \end{aligned} \tag{11}$$

where α, β, and γ are regularization parameters.

There are three variables {M, B, Y} needed to be tuned in the SRDML approach. An alternating optimization method is used to solve Eq. 11.

Tune **B**. With fixed **M** and **Y**, the objective function of **B** can be written as,

$$\begin{aligned} \min_B \ & \|\text{X-BY}\|_2^2, \\ s.t. \quad & \|b_i\|^2 = 1, \ \forall i, \end{aligned} \tag{12}$$

Obviously, it is a least square problem with quadratic constrains. The Lagrange dual function is used, and we have,

$$h(B, \tau) = \left(\|\text{X-BY}\|_2^2 + \sum_i^K \tau_i(\|b_i\|^2 - 1)\right), \tag{13}$$

where τ is the Lagrange vector. A diagonal matrix $\Delta \in R^{K\times K}$ is introduced, where $\Delta_{ii} = \tau_i$ for all *i*. We can obtain the following problem,

$$h(B, \tau) = \left(\|\text{X-BY}\|_2^2 + Tr(B^T B\Delta) - Tr(\Delta)\right). \tag{14}$$

Taking the first-order partial derivatives of **B** in Eq. 14, we can obtain the close-solution of **B** as,

$$B^* = XY^T(YY^T + \Delta)^{-1}. \tag{15}$$

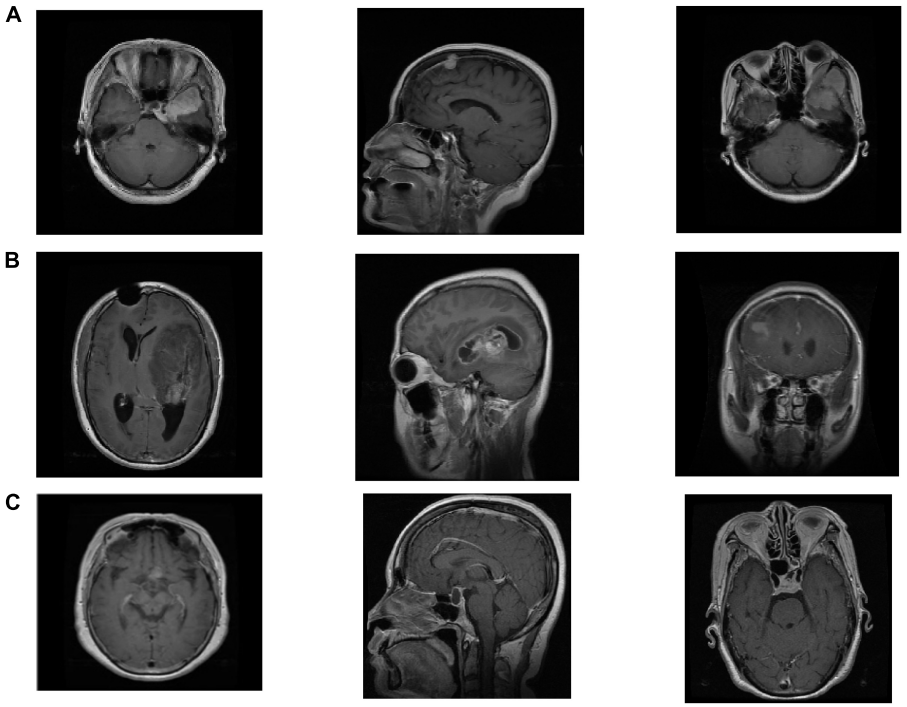


FIGURE 1 | The example samples in the Cheng dataset, **(A)** Meningioma, **(B)** Glioma, and **(C)** Pituitary.

Tune **Y**. With fixed **M** and **B**, the objective function of **Y** can be written as,

$$\min_{Y} \|\text{X-BY}\|_2^2 + \alpha Tr(\text{YLY}) + \beta \sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right) + \gamma \|Y\|_2^2, \tag{16}$$

Taking the first-order partial derivatives of y_i in Eq. 16, we can obtain the close-solution of y_i as,

$$y_i^* = (B^T B + \gamma I + \alpha L + \beta \sum_{i,j} g'(\Lambda) M)^{-1} (B^T Y + \beta M \sum g'(\Lambda) y_j), \tag{17}$$

where $\Lambda = 1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)$.

Tune **M**. With fixed **Y** and **B**, the objective function of **M** can be written as,

$$\min_{M} \beta \sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right), \quad s.t.\ M \geq 0, \tag{18}$$

TABLE 1 | Retrieval performance of SRDML for the query image Meningioma on 5-fold test set.

	*m*AP	prec@10	prec@20
Fold 1	92.12	89.93	89.93
Fold 2	92.13	89.95	89.93
Fold 3	92.14	89.92	89.88
Fold 4	92.19	89.97	89.96
Fold 5	92.19	89.95	89.92
Mean (standard deviation)	92.15 (0.030)	89.94 (0.017)	89.92 (0.026)

TABLE 2 | Retrieval performance of SRDML for the query image Glioma on 5-fold test set.

	*m*AP	prec@10	prec@20
Fold 1	97.87	95.75	95.74
Fold 2	97.86	95.71	95.72
Fold 3	97.81	95.68	95.70
Fold 4	97.80	95.70	95.70
Fold 5	97.75	95.68	95.67
Mean (standard deviation)	97.82 (0.044)	95.70 (0.026)	95.70 (0.023)

TABLE 3 | Retrieval performance of SRDML for the query image Pituitary on 5-fold test set.

	*m*AP	prec@10	prec@20
Fold 1	97.87	96.45	96.44
Fold 2	97.93	96.43	96.46
Fold 3	97.94	96.50	96.50
Fold 4	97.96	96.51	96.47
Fold 5	97.87	96.46	96.54
Mean (standard deviation)	97.92 (0.037)	96.47 (0.030)	96.48 (0.035)

The Lagrange dual function is used, and we have,

$$h(M, \lambda) = \beta \sum_{i,j} g\left(1 - \ell_{ij}\left(\tau - d_M^2(y_i - y_j)\right)\right) + \lambda \|\mathrm{M\text{-}I}\|_2^2, \tag{19}$$

where λ is the Lagrange parameter.

Let $s_{ij} = (y_i - y_j)^T (y_i - y_j)$, taking the first-order partial derivatives of **M** in Eq. 19, we obtain,

$$\frac{\partial h}{\partial M} = \beta \sum_{\{x_i, x_j\} \in A_{set}} g'(\Lambda) \mathrm{C}_{ij} + 2\gamma (M - I). \tag{20}$$

Then we can obtain M by $M_t = M_{t-1} - \xi \frac{\partial h}{\partial M_{t-1}}$, where ξ is the learning rate in gradient descent method.

The proposed SRDML approach is shown in algorithm 1.

Algorithm 1 SRDML approach
Input: Similar image pair subset and dissimilar image pair subset in **X**, parameters α, β, and γ;
Output: coding coefficient matrix **Y**, dictionary **B**, and metric matrix **M**;
1: Initializing **B** and **Y** with the K-SVD algorithm, **M** with the identity matrix;
2: Calculating graph Laplacian matrix **L** using Eqs 8 and 9;
3: while $t \leq$ maximum number of iterations T_{max}
  Calculating dictionary **B** using Eq. 15;
  Calculating graph Laplacian matrix **L** using Eqs 8 and 9;
  Calculating coding coefficient matrix **Y** using Eq. 17;
  Calculating metric matrix **M** using Eq. 20;
4: Obtaining $B^* = B^{T_{max}}$, $Y^* = Y^{T_{max}}$, and $M^* = M^{T_{max}}$.

EXPERIMENTS

Experimental Settings

We use the Cheng brain MRI dataset (Cheng et al., 2016) to evaluate the effectiveness of the proposed SRDML approach. This brain MRI dataset contains a total of 3064 slices from 233 patients. The brain MRI images contain three types of tumor images Meningioma, Glioma, and Pituitary. The numbers of the three types of tumor images are 708, 1426, and 930, respectively. The size of each image is 512 × 512 pixels. 600 images of each type of brain tumor are selected in the experiment. **Figure 1** depicts the example samples of the Cheng dataset.

We use the GIST descriptor implemented by Gumaei et al. (2018, 2019) in this study. GIST descriptor extracts the image features based on the spatial envelope. In the experiment, we obtain 512 dimensional features of each image. We randomly divide brain MRI images into five subgroups of the same size. It is ensured that the different types of tumors in the five subgroups do not overlap and had equal proportions. We use 5-fold cross-validation to evaluate performance. Each image in the test data set is regarded as a query image. We adopt the mean average precision (*mAP*) and top-n retrieval precision (named as Prec@n) (Cheng et al., 2016). *mAP* and Prec@n can be calculated as

$$mAP = \frac{\sum_{i=1}^{n} \frac{i}{rank(i)}}{n}, \tag{21}$$

where n represents the number of retrieved images of the same type of brain tumor in the dataset, $rank(i)$ represents the ranking number of the i-th retrieved image of the same type of brain tumor in the search results.

$$Prec@n = \frac{1}{s} \sum_{i=1}^{s} \sum_{j=1}^{k} \frac{j}{Position(j)}, \tag{22}$$

where s is the number of queries, and $Position(j)$ refers to the position of the j-th relevant sample in the search results.

The retrieval performance of SRDML is compared with various matrix learning algorithms LMNN (Weinberger and Saul, 2009), RCA (Bar-Hillel et al., 2005), MPP (Yang et al., 2012), TFFHD (Sun et al., 2018), and RDML-CCPVL (Ni et al., 2018). The parameter setting of the comparison algorithm refers to

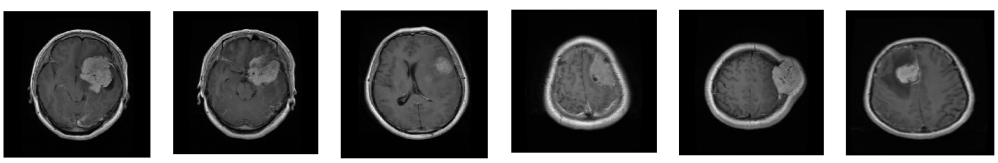


FIGURE 2 | Retrieval result examples of SRDML in Meningioma.

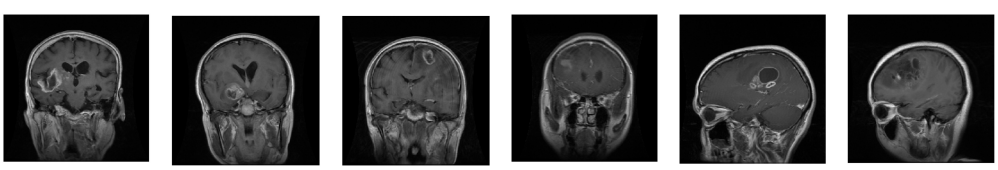


FIGURE 3 | Retrieval result examples of SRDML in Glioma.

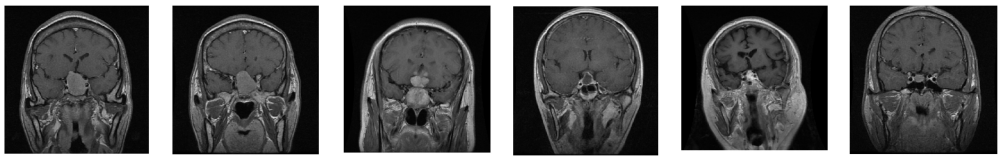


FIGURE 4 | Retrieval result examples of SRDML in Pituitary.

the setting of the original literature. The parameters of SRDML are set as follows. The size of the dictionary is equal to the number of the training set. The regularization parameters are set in the grid $\{10^{-3}, 10^{-2}, ..., 10\}$. The dimension of metric matrix is set in the grid {50, 100, 200, ..., 500}. The sharpness parameter in generalized logistic loss function is set to be 1. The KNN parameter is set in the grid {3, 5, ..., 13}. The running environment of all algorithms is CPU i7-8700k, 3.2 GHZ, 32GB RAM, and the running software is Matlab 2019.

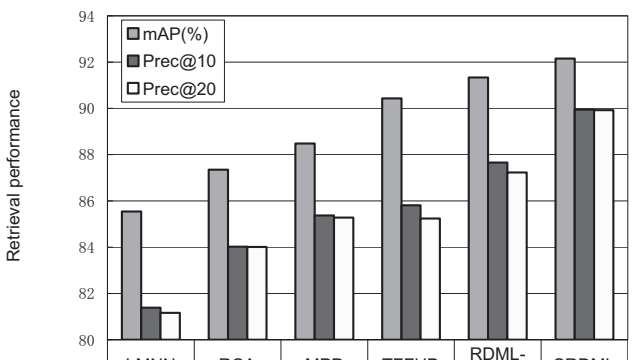
Performance Evaluation

Tables 1–**3** list the retrieval performance of SRDML for the query image Meningioma, Glioma, Pituitary, respectively. These tables show the results of the 5-fold cross validation test set. We can see that the results obtained by the SRDML approach are consistent on the 5-fold test set, which verifies that the retrieval performance of our approach is stable. We conduct further case analysis on the SRDML. **Figures 2**–**4** depict the retrieval result examples of SRDML in Meningioma, Glioma, Pituitary, respectively. The first column of images in **Figures 2**–**4** are the query images. The second to sixth images are five retrieved images. We can see that the retrieved images are all highly related to the query images. The results show that SRDML can give full play to its advantages in brain MRI sets. The SRDML approach uses sparse representation and metric learning to mine the structural features and discrimination information of the data. The local information term and similarity constraints based on coding coefficients ensure the "maximum inter-class dispersion" and "minimum intra-class dispersion" of the extracted brain MRI features. Thus, the SRDML approach has the important practical value in assisting clinical diagnosis.

Performance Comparison

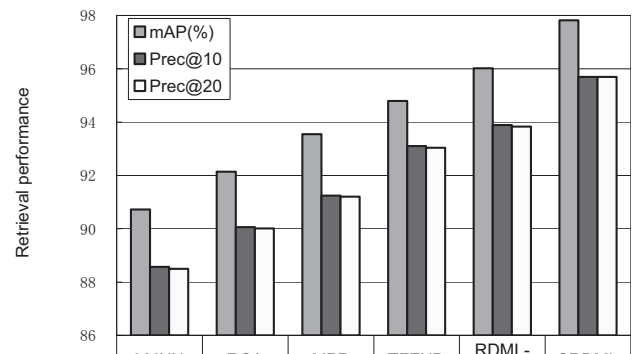
We use the SRDML approach to compare the performance with LMNN, RCA, MPP, TFFHD, and RDML-CCPVL. **Figures 5**–**7** show the retrieval performance of various comparison algorithms in Meningioma, Glioma, and Pituitary. The results show these methods in retrieval performance of *m*AP, Prec@10, as well as Prec@20. According to the results in **Figures 5**–**7**, the following results are obtained:

(1) In the Meningioma, the SRDML approach is compared with the second-best in *m*AP, Prec@10, as well as Prec@20 by 0.81, 1.28, and 2.69%, respectively. SRDML has excellent retrieval performance, especially in Prec@20. This shows that the proposed approach has high retrieval performance. In the Glioma, the SRDML approach is compared with the second-best approach in *m*AP, Prec@10 as well as Prec@20 increased by 1.80, 1.81, and 1.87%, respectively. In the Pituitary, the SRDML approach is compared with the second-best approach in *m*AP, Prec@10, as well as Prec@20 increased by 1.37, 2.20, and 2.26%, respectively.
(2) The SRDML approach has significantly improved the retrieval performance compared with the traditional



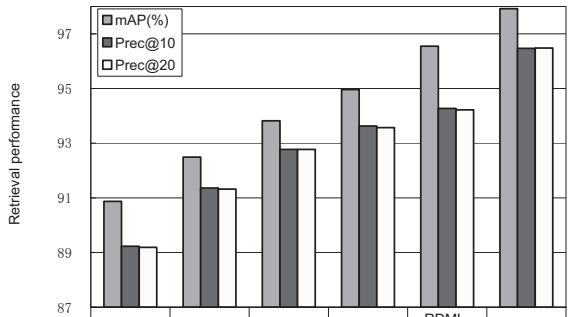
	LMNN	RCA	MPP	TFFHD	RDML-CCPVL	SRDML
mAP(%)	85.54	87.35	88.48	90.43	91.34	92.15
Prec@10	81.38	84.02	85.37	85.81	87.66	89.94
Prec@20	81.16	84.01	85.28	85.24	87.23	89.92

FIGURE 5 | Retrieval performance comparisons in Meningioma.



	LMNN	RCA	MPP	TFFHD	RDML-CCPVL	SRDML
mAP(%)	90.72	92.14	93.55	94.79	96.02	97.82
Prec@10	88.57	90.06	91.24	93.10	93.89	95.70
Prec@20	88.50	90.01	91.20	93.04	93.83	95.70

FIGURE 6 | Retrieval performance comparisons in Glioma.



	LMNN	RCA	MPP	TFFHD	RDML-CCPVL	SRDML
mAP(%)	90.87	92.49	93.82	94.96	96.55	97.92
Prec@10	89.23	91.36	92.77	93.63	94.27	96.47
Prec@20	89.19	91.32	92.77	93.57	94.22	96.48

FIGURE 7 | Retrieval performance comparisons in Pituitary.

TABLE 4 | *m*AP(%) of the SRDML approach with different *k*.

	$k = 3$	$k = 5$	$k = 7$	$k = 9$	$k = 11$	$k = 13$
Meningioma	90.53	91.77	**92.15**	**92.15**	92.00	91.42
Glioma	96.50	97.08	97.69	**97.82**	97.47	96.99
Pituitary	97.09	97.51	**97.92**	97.89	97.44	97.01

The bold values represent the best results in comparison experiments.

TABLE 5 | *m*AP(%) of the SRDML approach with different *m*.

	$m = 50$	$m = 100$	$m = 200$	$m = 300$	$m = 400$	$m = 500$
Meningioma	89.98	91.43	**92.15**	92.10	92.12	91.58
Glioma	96.90	**97.82**	**97.82**	97.59	97.60	97.01
Pituitary	96.86	**97.92**	97.90	97.80	97.33	97.02

The bold values represent the best results in comparison experiments.

approaches. The SRDML approach uses the sparse representation to learn the robust feature representation of brain MRI images, and learns the metric space with discrimination ability, so that the similar brain MRI pairs are closely mapped in the metric space and the dissimilar brain MRI pairs are separated from each other as much as possible.

Parameter Analysis

The parameters of the SRDML approach are obtained by the grid optimization method. Here we mainly evaluate the influence of KNN parameter k and the dimension of metric matrix m. **Table 4** lists the *m*AP(%) of the SRDML approach with different k. The parameter k is set in the grid {3, 5, …, 13}. The value of k increases from 3 to 7, and the *m*AP value increases significantly. It is verified that the local information item is an important part in the SRDML approach. Extracting the appropriate local information can be helpful to describe the image features and improve retrieval accuracy. When the value of k increases from 9 to 13, the retrieval performance begins to decrease, indicating that too large the value of k will cause useless information of the local information and too much useless information, which affects the retrieval performance of SRDML. **Table 5** lists the *m*AP(%) of the SRDML approach with different m. The parameter m is set in the grid {50, 100, 200, …, 500}. It can be seen from the results in **Table 5** that when m is set as 100 or 200, *m*AP(%) of the SRDML approach can reach the optimal retrieval performance. If the value of m is too large or too small, the metric space cannot properly represent the internal data structure of brain MRI images. Therefore, it is feasible to use the grid search method to determine the optimal values of k and m.

CONCLUSION

The purpose of medical image retrieval is to retrieve similar image data from a huge imaging database. The retrieval results should not only be similar in image feature measurement, but also fit in image semantics as much as possible, to help doctors retrieve images with the same pathology. This paper proposes a similarity retrieval approach for brain MRI images based on sparse representation and metric learning. The characteristics of this approach are using sparse representation to extract the robust features of brain MRI image, which can more effectively represent the information of brain MRI image. The relative similarity constraint and local preserving information are used to project the low dimensional metric matrix, which improves the retrieval accuracy. The experiments are carried out on public data sets. The experiments show the effectiveness and accuracy of our approach in brain MRI image retrieval. The future work will be considered from the following aspects: (1) while improving the retrieval accuracy of brain MRI images, ensure the retrieval speed, to further provide assistance and support for doctors in diagnosing brain tumor lesions. (2) The experimental data scale of this paper is not large enough. During the retrieval process, several images with the highest similarity are returned by comparing the feature vector similarity of the data set in turn. When dealing with large-scale data, we will consider how to combine the technologies such as hash sorting of retrieval results to realize the real-time performance of retrieval results. (3) For different images and images with different principles, the feature algorithms involved are different. Forming a more general medical image retrieval system is the next research direction.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The Cheng dataset analyzed for this study can be found in this link [https://figshare.com/articles/dataset/brain_tumor_dataset/1512427].

AUTHOR CONTRIBUTIONS

GZ and TN conceived and designed the proposed model and wrote the manuscript. BL and XH ran the experiment and analyzed the results. All authors read and approved the manuscript.

FUNDING

This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK 20211333, the Science and Technology Project of Changzhou City under Grant CE20215032, and Jiangsu Future Network Scientific Research Fund Project under Grant FNSRFP-2021-YB-36.

REFERENCES

Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D., and Ridgeway, G. (2005). Learning a Mahalanobis metric from equivalence constraints. *J. Mach. Learn. Res.* 612, 937–965. doi: 10.1115/1.1902843

Cao, X., Zhang, H., Guo, X., Liu, S., and Meng, D. (2015). SLED: semantic label embedding dictionary representation for multilabel image annotation. *IEEE Trans. Image Process.* 24, 2746–2759. doi: 10.1109/TIP.2015.2428055

Chahal, P. K., Pandey, S., and Goel, S. (2020). A survey on brain tumor detection techniques for MR images. *Multimed. Tools Appl.* 79, 21771–21814. doi: 10.1007/s11042-020-08898-3

Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., et al. (2016). Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PLoS One* 11:e0157112. doi: 10.1371/journal.pone.0157112

Gao, S., Chia, L., Tsang, I. W., and Ren, Z. (2014). Concurrent single-label image classification and annotation via efficient multi-layer group sparse coding. *IEEE Trans. Multimedia* 16, 762–771. doi: 10.1109/TMM.2014.2299516

Gu, D., Liang, C., and Zhao, H. (2017). A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis. *Artif. Intell. Med.* 77, 31–47. doi: 10.1016/j.artmed.2017.02.003

Gu, X., Zhang, C., and Ni, T. (2021a). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1679–1687. doi: 10.1109/TCBB.2020.3006699

Gu, X., Shen, Z., Xue, J., Fan, Y., and Ni, T. (2021b). Brain tumor MR image classification using convolutional dictionary learning with local constraint. *Front. Neurosci.* 15:679847. doi: 10.3389/fnins.2021.679847

Gumaei, A., Sammouda, R., Malik, A., Al-Salman, S., and Alsanad, A. (2018). An improved multispectral palmprint recognition system using autoencoder with regularized extreme learning machine. *Comput. Intell. Neurosci.* 2018:8041609. doi: 10.1155/2018/8041609

Gumaei, A., Sammouda, R., Malik, A., Al-Salman, S., and Alsanad, A. (2019). Anti-spoofing cloud-based multi-spectral biometric identification system for enterprise security and privacy-preservation. *J. Parallel Distrib. Comput.* 124, 27–40. doi: 10.1016/j.jpdc.2018.10.005

Huang, M., Yang, W., Wu, Y., Jiang, J., Gao, Y., Chen, Y., et al. (2014). Content-based image retrieval using spatial layout information in brain tumor T1-weighted contrast-enhanced MR images. *PLoS One*. 9:e102754. doi: 10.1371/journal.pone.0102754

Huang, M., Yang, W., Yu, M., Lu, Z., Feng, Q., and Chen, W. (2012). Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images. *Comput. Math. Methods Med.* 2012:280538. doi: 10.1155/2012/280538

Jiang, Y., Gu, X., Wu, D., Hang, W., Xue, J., Qiu, S., et al. (2021). A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 40–52. doi: 10.1109/TCBB.2019.2963873

Lu, Z., Han, P., Wang, L., and Wen, J. (2015). Semantic sparse recoding of visual content for image applications. *IEEE Trans. Image Process.* 24, 176–188. doi: 10.1109/TIP.2014.2375641

Müller, H., and Unay, D. (2017). Retrieval from and understanding of large-scale multi-modal medical datasets: a review. *IEEE Trans. Multimed.* 19, 2093–2104. doi: 10.1109/TMM.2017.2729400

Nguyen, B., Morell, C., and Baets, B. D. (2020). Scalable large-margin distance metric learning using stochastic gradient descent. *IEEE Trans. Cybern.* 50, 1072–1083. doi: 10.1109/TCYB.2018.2881417

Ni, T., Ding, Z., Chen, F., and Wang, H. (2018). Relative distance metric leaning based on clustering centralization and projection vectors learning for person re-identification. *IEEE Access* 6, 11405–11411. doi: 10.1109/ACCESS.2018.2795020

Srinivas, U., Mousavi, H. S., Monga, V., Hattel, A., and Jayarao, B. (2014). Simultaneous sparsity model for histopathological image representation and classification. *IEEE Trans. Med. Imaging* 33, 1163–1179. doi: 10.1109/TMI.2014.2306173

Sun, X., Zhang, N., Wu, H., Yu, X., Wu, X., and Yu, S. (2018). Medical image retrieval approach by texture features fusion based on Hausdorff distance. *Math. Probl. Eng.* 2018:7308328. doi: 10.1155/2018/7308328

Tang, J. H., Hong, R. C., Yan, S. C., and Chua, T. S. (2010). Image annotation by kNN-sparse graph-based label propagation over noisily-tagged Web images. *ACM Trans. Intell. Syst. Technol.* 2, 1–15. doi: 10.1145/1899412.1899418

Verma, Y., and Jawahar, C. V. (2017). Image annotation by propagating labels from semantic neighbourhoods. *Int. J. Comput. Vis.* 121, 126–148. doi: 10.1007/s11263-016-0927-0

Weinberger, K. Q., and Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* 10, 207–244. doi: 10.1007/s10845-008-0108-2

Yang, W., Feng, Q., Yu, M., Lu, Z., Gao, Y., Xu, Y., et al. (2012). Content-based retrieval of brain tumor in contrast-enhanced MRI images using tumor margin information and learned distance metric. *Med. Phys.* 39, 6929–6942. doi: 10.1118/1.4754305

Zhang, X., Dou, H., Ju, T., Xu, J., and Zhang, S. (2015). Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis. *IEEE J. Biomed. Health Inform.* 20, 1377–1383. doi: 10.1109/JBHI.2015.2461671

Zhang, Y., Chen, C., Tian, Z., Feng, R., Cheng, Y., and Xu, J. (2019). The diagnostic value of MRI-based texture analysis in discrimination of tumors located in posterior fossa: a preliminary study. *Front. Neurosci.* 13:1113. doi: 10.3389/fnins.2019.01113

Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Ma, Y., Shi, H., et al. (2018). Histopathological whole slide image analysis using context-based CBIR. *IEEE Trans. Med. Imaging* 37, 1641–1652. doi: 10.1109/TMI.2018.2796130

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 20 January 2022
doi: 10.3389/fnins.2021.694010

Continuous Monitoring of Mental Load During Virtual Simulator Training for Laparoscopic Surgery Reflects Laparoscopic Dexterity: A Comparative Study Using a Novel Wireless Device

Neta B. Maimon[1,2†], Maxim Bez[3†], Denis Drobot[4,5], Lior Molcho[2], Nathan Intrator[2,6], Eli Kakiashvilli[5] and Amitai Bickel[4,5]*

[1] The School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel, [2] Neurosteer Ltd, Herzliya, Israel, [3] Medical Corps, Israel Defense Forces, Ramat Gan, Israel, [4] Faculty of Medicine in the Galilee, Bar-Ilan University, Safed, Israel, [5] Department of Surgery A, Galilee Medical Center, Nahariyya, Israel, [6] Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Ji Won Bang, NYU Grossman School of Medicine, United States
Yufeng Yao, Changshu Institute of Technology, China

***Correspondence:**
Neta B. Maimon
netacoh3@mail.tau.ac.il

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

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *12 April 2021*
Accepted: *22 November 2021*
Published: *20 January 2022*

Citation:
Maimon NB, Bez M, Drobot D, Molcho L, Intrator N, Kakiashvilli E and Bickel A (2022) Continuous Monitoring of Mental Load During Virtual Simulator Training for Laparoscopic Surgery Reflects Laparoscopic Dexterity: A Comparative Study Using a Novel Wireless Device. Front. Neurosci. 15:694010. doi: 10.3389/fnins.2021.694010

Introduction: Cognitive Load Theory (CLT) relates to the efficiency with which individuals manipulate the limited capacity of working memory load. Repeated training generally results in individual performance increase and cognitive load decrease, as measured by both behavioral and neuroimaging methods. One of the known biomarkers for cognitive load is frontal theta band, measured by an EEG. Simulation-based training is an effective tool for acquiring practical skills, specifically to train new surgeons in a controlled and hazard-free environment. Measuring the cognitive load of young surgeons undergoing such training can help to determine whether they are ready to take part in a real surgery. In this study, we measured the performance of medical students and interns in a surgery simulator, while their brain activity was monitored by a single-channel EEG.

Methods: A total of 38 medical students and interns were divided into three groups and underwent three experiments examining their behavioral performances. The participants were performing a task while being monitored by the Simbionix LAP MENTOR™. Their brain activity was simultaneously measured using a single-channel EEG with novel signal processing (Aurora by Neurosteer®). Each experiment included three trials of a simulator task performed with laparoscopic hands. The time retention between the tasks was different in each experiment, in order to examine changes in performance and cognitive load biomarkers that occurred during the task or as a result of nighttime sleep consolidation.

Results: The participants' behavioral performance improved with trial repetition in all three experiments. In Experiments 1 and 2, delta band and the novel VC9 biomarker (previously shown to correlate with cognitive load) exhibited a significant decrease in

activity with trial repetition. Additionally, delta, VC9, and, to some extent, theta activity decreased with better individual performance.

Discussion: In correspondence with previous research, EEG markers delta, VC9, and theta (partially) decreased with lower cognitive load and higher performance; the novel biomarker, VC9, showed higher sensitivity to lower cognitive load levels. Together, these measurements may be used for the neuroimaging assessment of cognitive load while performing simulator laparoscopic tasks. This can potentially be expanded to evaluate the efficacy of different medical simulations to provide more efficient training to medical staff and measure cognitive and mental loads in real laparoscopic surgeries.

Keywords: cognitive load, surgical simulator training, EEG biomarker, laparoscopic operations, brain assessment, mental load assessment, machine learning

INTRODUCTION

Medical training of surgical interns in recent years utilizes digital simulators as an effective tool for acquiring practical skills in a controlled and hazard-free environment. These simulators help train interns and provide their performance outputs for each task, thus supporting the learning process. The surgery simulators offer practice of various tasks and present an evaluation of the performance *via* behavioral measurements (i.e., performance accuracy and performance length, etc.). Nevertheless, the simulator output measurements have two main drawbacks. First, althoughthe behavioral performance serves as a good indication of laparoscopic dexterity, it does not give an indication regarding the student's cognitive state and, specifically, his or her cognitive load level during the task performance. Assessing cognitive load and stress levels directly during the surgery simulation is essential to evaluate the intern's ability to engage in real-life operations. Second, the behavioral performance is measured while completing a task under the surgery simulator. Therefore, it does not allow extraction of similar measurements during laparoscopic procedures inside the operating room. To address these points, we designed three experiments in which medical students and interns perform laparoscopic tasks under a surgery simulator, while recording their brain activity with a wearable single-channel EEG headset. The experiments measured behavioral performance, EEG bands that are known to correlate to cognitive load, and a novel machine-learning-based cognitive load biomarker. In the following paragraphs, we will review the literature regarding cognitive load, EEG biomarkers of cognitive load, medical simulators, and previous measurements of cognitive load during simulator tasks.

Cognitive Load

Any acquirement of a new skill or knowledge must be passed through working memory (WM) before being transferred into long-term memory (LTM). According to Cognitive Load Theory (CLT, Miller, 1956), WM resources, unlike LTM resources, are limited in their capacity for processing or holding new information. However, when performing a complex task, the new information elements are being processed simultaneously and not iterated, so their interactions cause a much higher WM load. According to CLT, optimizing learning processes may be achieved by constructing automated schemas in which the WM load is reduced (Jolles et al., 2013). The amount of cognitive load used is a good predictor of the learning process during a task performance, and continuous measurement is of particular importance (Van Merriënboer and Sweller, 2005; Constantinidis and Klingberg, 2016). For instance, practicing the same task repeatedly will cause some of the interacting information elements to be stored together in a schematic form in LTM, so that they can be extracted and manipulated in WM as single-information elements (Schneider and Shiffrin, 1977).

Cognitive Load Biomarkers

Several methods were previously described and validated as measures of WM and cognitive load. Traditionally, subjective self-rating scales were proven as reliable assessment tools across several studies (Paas et al., 2003). However, as this tool can only be recorded crudely and retrospectively, objective assessment methods with a real-time indication of WM are in demand. By using such an approach, the activity can be broken down into different components that reflect different stages of complex simulations and be used to evaluate the efficacy of each training element. Several biological methods have been reported to successfully assess WM, such as pupil size (Van Gerven et al., 2004), tracked eye movement (Van Gog and Scheiter, 2010), salivary cortisol levels (Bong et al., 2010), and functional magnetic resonance imaging (fMRI) (Van Dillen et al., 2009).

Using electroencephalography (EEG), studies repeatedly show that the theta band (4–7 Hz) measured by frontal electrodes increases with higher cognitive load and task difficulty (Scheeringa et al., 2009; Antonenko et al., 2010). Multiple studies confirmed this correspondence using a variety of cognitive tasks, including the *n*-back (Gevins et al., 1997), operation span task (Scharinger et al., 2017), visuospatial WM tasks (Bastiaansen et al., 2002), and simulated real-life experiences such as in driving (Di Flumeri et al., 2018) or flight simulators (Dussault et al., 2005).

In addition to theta, delta oscillation (0.5–4 Hz) was also found to increase with higher cognitive load and/or task difficulty such as during reading tasks (Zarjam et al., 2011) and multiple-choice reaction tasks (Schapkin et al., 2020), and to increase with higher vigilance during low-cognitive-load attention tasks (Kim et al., 2017). It has been suggested that delta activity is an indicator of internal attention, and therefore increases while undergoing

mental tasks (Harmony et al., 1996). Delta activity was also associated with interference inhibition processes, which occur in order to modulate sensory afferences, which in turn increase internal concentration (Harmony, 2013).

Medical Simulations

Medical simulations are a common and widespread tool for medical education (Gurusamy et al., 2009). Medical simulations can emulate common scenarios in clinical practice, and through interactive interplay and hands-on teaching, they can improve the effectiveness and quality of teaching of healthcare professionals (Kunkler, 2006). These simulations can be particularly beneficial for surgical staff, as they allow residents to practice and perfect complex procedures, ensuring they have enough experience and practice before real patient contact (Thompson et al., 2014). As laparoscopic surgeries demand unique eye–hand coordination and are performed while the surgeon indirectly observes the intra-abdominal contents without tactile sensation but through a camera view, these surgeons are ideal candidates for virtual reality (VR) simulators. Indeed, these simulators have been shown to greatly improve surgical operating skills and reduce operating time (Gallagher et al., 2005; Larsen et al., 2009; Gauger et al., 2010; Yiannakopoulou et al., 2015; Tucker et al., 2017).

Recently, a few studies made the connection between CLT and medical training using surgery simulators (Andersen et al., 2016, 2018; Frederiksen et al., 2020). In these studies, participants completed laparoscopic tasks *via* VR simulators, while their WM loads were estimated using a secondary behavioral task (e.g., reaction times in response to auditory stimuli). They found that different interventions with medical VR training may modulate WM load. For example, Frederiksen et al. (2020) showed that immersive VR results show higher cognitive load than the conventional VR training, and Andersen et al. (2018) demonstrated that additional VR training sessions may reduce cognitive load and increase the performance of dissection training. However, these studies evaluated cognitive load to assess the effect of an external manipulation, rather than to understand the neurophysiological mechanism underlying the simulation process itself. None of the studies recorded an objective physiological WM load biomarker such as pre-frontal cortex activity or frontal theta oscillation during the training itself. Such measurements would not only exhibit a continuous level of cognitive load, but, if found reliable, could later be measured in real scenarios. Lower levels of WM load, along with efficient performance, may indicate the readiness levels of surgeons. Additionally, a wearable, portable device may enable real-time monitoring of WM load levels inside the operating room, reflecting a surgeon's abilities and laparoscopic dexterity.

The Present Study

The aim of this study was to explore the neural mechanism that underlies surgery simulation training. Specifically, our aim was to test the relationship between cognitive load, skill acquisition, and the activity levels of different brain oscillations. We aimed to track cognitive load neuro-markers using an EEG device that will enable hand mobility while performing surgical simulator tasks. Importantly, we wanted to test medical students and interns who, on the one hand, have great motivation and motor/cognitive abilities to perform such laparoscopic tasks and, on the other, have no previous hands-on experience, neither in surgery simulators nor with real-life patients. Finally, we intended to compare "online gains"—improvements in performance that occur while undergoing the task—to "offline gains," the improvements in skill acquisition preceding a consolidation during nighttime sleep between task trainings, without further practice (Issenberg et al., 2011; Fraser et al., 2015; Lugassy et al., 2018). Notwithstanding, frontal theta and other known cognitive load biomarkers were mostly used to measure WM load during task completion itself and not to evaluate the load differences between tasks' sessions.

To meet these goals, we divided 38 medical students and interns into three experiments. Each experiment included three trials of the same short laparoscopic task administered by a surgical simulator (Simbionix LAP MENTOR^{TM}). While performing the tasks, participants' brain activity was recorded using a wearable single-channel EEG device (Aurora by Neurosteer® Inc), from which we continuously measured frontal brain oscillations (i.e., the theta and delta bands) throughout the simulation task. In each experiment, the participants repeated the same task three times, thus allowing them to obtain the new skill and decrease their cognitive load levels.

In addition to known brain oscillations, Neurosteer® also provided us with a machine-learning-based cognitive load biomarker (i.e., VC9). VC9 activity was previously shown to increase with cognitive load in the standard and most common WM tasks (i.e., the *n*-back task, Maimon et al., 2020), auditory recognition and classification tasks (Molcho et al., 2021), and interruptions paradigm (Bolton et al., 2021). Specifically, Maimon et al. (2021) showed that the VC9 biomarker is more sensitive to the finer differences between cognitive loads. During the *n*-back task, while both VC9 activity and theta and delta oscillations increased with higher cognitive loads, the VC9 biomarker exhibited higher sensitivity than the theta and delta oscillations to the lower levels of cognitive load. These subtler differences are particularly crucial for the present study's purposes and experimental design. With each task repetition, participants' cognitive load levels descended, eventually reaching relatively low levels of cognitive load. Therefore, the ability to detect finer differences between the low cognitive loads is the most critical for estimating participants' readiness to go into the operating room.

Thus, our hypotheses were as follows: (1) Behavioral performance will improve with repetition of the surgery simulator task. (2) This improvement of behavioral performance with task repetition will result in the decrease of cognitive load that will modulate the EEG features (i.e., theta, delta, and VC9 will decrease with task repetition). (3) EEG features will correlate to some extent with higher individual performance (i.e., decrease with better individual performance), reflecting the reduced need for cognitive resources together with improving laparoscopic dexterity. (4) "Offline gains" will be present following procedural memory consolidation during nighttime sleep. Probing these hypotheses will help reveal new and objective information regarding the efficacy of simulation-based training. For a graphical representation

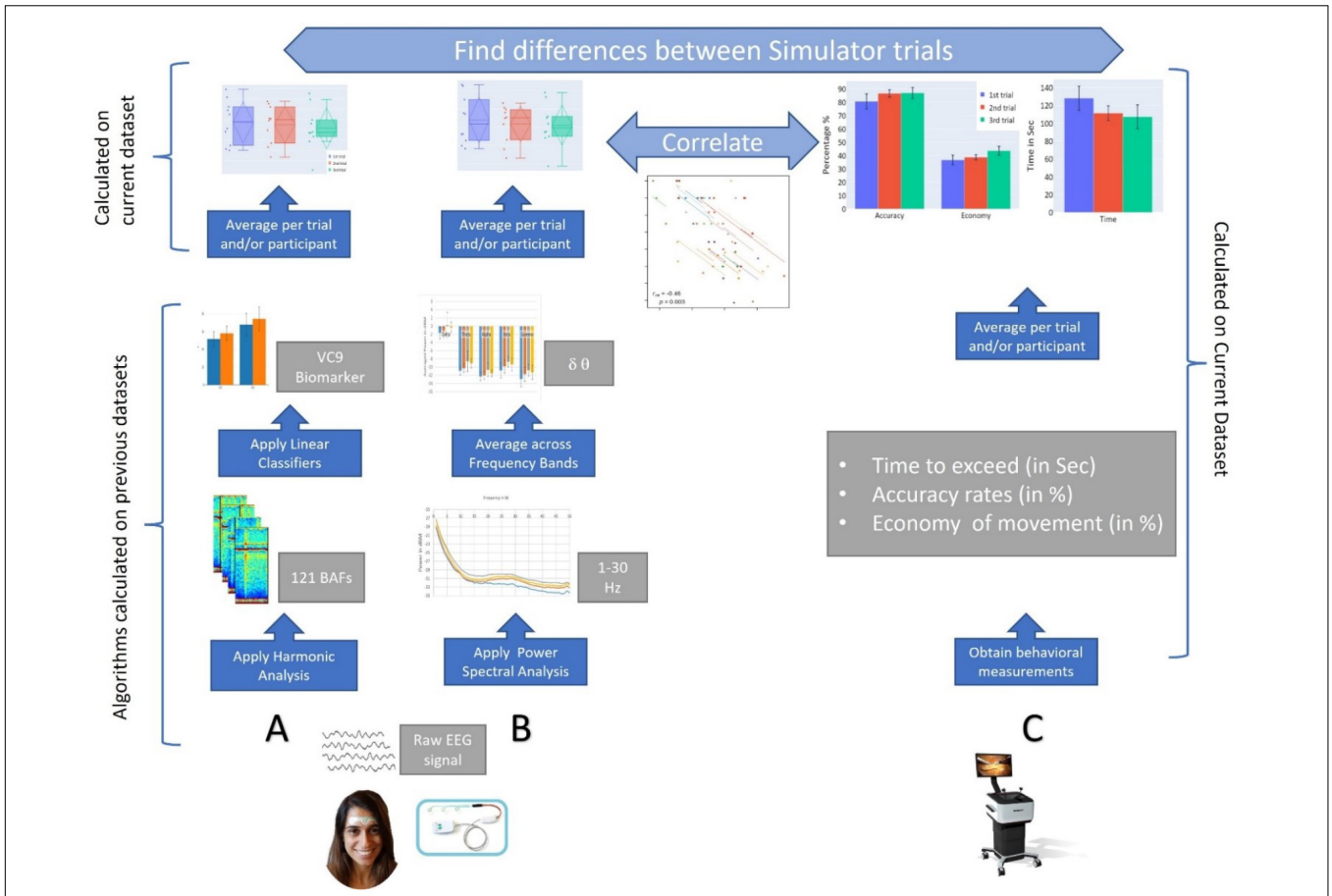


FIGURE 1 | A graphical representation of the study. **(A)** EEG data are recorded *via* a single channel using a three-electrode forehead patch by an Aurora EEG system while participants perform the simulator task. Data are pre-processed by a predefined set of wavelet packet atoms to produce 121 Brain Activity Features (BAFs). The BAFs are passed through linear models, which were pretrained on external data, to form a higher-level biomarker, VC9. Then, VC9 activity is averaged per trial and participant to find differences between simulator trials. **(B)** From the raw EEG data, power spectral density analysis is applied to find energy representation (in dB) produced for 1–50 Hz. The power density per Hz is averaged over frequency bands. Theta and delta activity are averaged per trial and participant to find differences between simulator trials. **(C)** Behavioral data are obtained from the Simbionix LAP MENTOR™ (i.e., time, accuracy, and economy of movement). These are averaged per trial and participant to find differences between simulator trials. Finally, correlations are calculated between EEG features (i.e., VC9, theta, and delta bands) and behavioral performance (time, accuracy, and economy of movement).

of the data analysis applied in the present study, see **Figure 1**.

EXPERIMENT 1

Materials and Methods

Sample Size Selection

To estimate the sample size required to observe a significant effect of simulator trial repetition in the present study, we conducted an *a priori* sample size analysis with G*Power software (Erdfelder et al., 1996). We based this analysis on the study by Andersen et al. (2016). Their study examined 18 novice medical students who underwent VR simulation training while their WM loads were examined using a secondary response time (RT) task. The experimental design included two groups: a control group ($n = 9$), which received standard instructions, and an intervention group ($n = 9$), which received CLT-based instructions. Similar to our present study, both groups in the Andersen et al. (2016) study underwent two post-training virtual procedures. Therefore, we based our power analysis on the within-group comparison of RTs between the first and second simulator trials of the control group. Moreover, Andersen et al. (2016) inserted these multiple within-participants fix effects into mixed linear models design, which was applied in the current study, as well; they are therefore most suitable for comparison. We conducted this analysis with a desired alpha of 0.05, and a power of 0.80, and calculated effect size from the control group's average RTs and standard deviation (SD) difference between the two simulator trials. The output of the analysis software was that the minimum required sample size was nine participants.

Participants

Nineteen participants (68% females, mean age 28, age range 25–36) were enrolled in the first experiment. All participants

were healthy medical interns who had completed 6 years of medical studies, never participated in laparoscopic surgery, and had no prior experience using a surgical simulator. Ethical approval for this study was granted by the Galilee Medical Center Institutional Review Board.

Electroencephalography Device

The EEG signal acquisition system included a three-electrode patch attached to the subject's forehead (Aurora by Neurosteer®., Herzliya, Israel). The medical-grade electrode patch included dry gel for optimal signal transduction. The electrodes were located at Fp1 and Fp2, with a reference electrode at Fpz. The EEG signal was amplified by a factor of 100 and sampled at 500 Hz. Signal processing was performed automatically by Neurosteer® in their cloud (see section "Signal Processing" below and **Supplementary Appendix A**). We were therefore provided with a sample per second of activity of the brain oscillations (i.e., delta, theta, alpha, beta, and gamma) and the VC9 biomarker.

Signal Processing

Full technical specifications regarding the signal analysis are provided in **Supplementary Appendix A**. In brief, the signal-processing algorithm interprets the EEG data using a time/frequency wavelet-packet analysis, instead of the commonly used spectral analysis. The Neurosteer® signal-processing algorithm interprets the EEG data using a time/frequency wavelet-packet analysis, creating a presentation of 121 brain activity features (BAFs) composed of the fundamental frequencies and their high harmonics.

To demonstrate this process, let g and h be a set of biorthogonal quadrature filters created from the filters G and H, respectively. These are convolution-decimation operators, where, in a simple Haar wavelet, g is a set of averages and h is a set of differences.

Let ψ_1 be the mother wavelet associated with the filters $s \in H$, and $d \in G$. Then, the collection of wavelet packets ψ_n, is given by:

$$\psi_{2n} = H\psi_n;\ \psi_{2n}(t) = \sqrt{2}\sum_{j\in Z} \mathrm{s}(j)\,\psi_n(2t-j), \quad (1)$$

$$\psi_{2n+1} = G\psi_n;\ \psi_{2n+1}(t) = \sqrt{2}\sum_{j\in Z} \mathrm{d}(j)\,\psi_n(2t-j). \quad (2)$$

The recursive form provides a natural arrangement in the form of a binary tree. The functions ψ_n have a fixed scale. A library of wavelet packets of any scale s, frequency f, and position p is given by:

$$\psi_{sfp}(t) = 2^{-s/2}\psi_f\left(2^{-s}t-p\right). \quad (3)$$

The wavelet packets $\{\psi_{sfp} : p \in Z\}$ include a large collection of potential orthonormal bases. An optimal basis can be chosen by the best-basis algorithm (Coifman and Wickerhauser, 1992). Furthermore, an optimal mother wavelet can be chosen by Neretti and Intrator (2002). Following robust statistics methods to prune some of the basis functions, one gets 121 basis functions, which we term brain activity features (BAFs). Based on a given labeled-BAFs dataset, various models can be created for different discriminations of these labels. In the linear case, these models are of the form:

$$V_k(w,x) = \Psi\left(\sum_i w_i x_i\right), \quad (4)$$

where, w is a vector of weights and Ψ is a transfer function that can either be linear, e.g., $\Psi(y) = y$, or sigmoidal for logistic regression $\Psi(y) = 1/(1+e^{-y})$.

From these BAFs, several linear and non-linear combinations were obtained using machine learning techniques on previously collected, labeled datasets. These datasets included data about participants undergoing different cognitive and emotional tasks. Specifically, VC9 was found with linear discriminant analysis technique (LDA, Ye and Li, 2005) on the 121 BAFs to be the best separator between an auditory detection task (higher cognitive load) and an auditory classification task (lower cognitive load, Molcho et al., 2021). VC9 was also recently validated as a cognitive load biomarker using the *n*-back task, auditory detection task, and interruption task (Maimon et al., 2020; Bolton et al., 2021; Molcho et al., 2021). It was shown that VC9 activity increased with increasing levels of cognitive load within cognitively healthy participants (Maimon et al., 2020; Bolton et al., 2021), and showed higher sensitivity than the theta band (Maimon et al., 2021). Neurosteer® provided a sample per second activity of EEG bands (i.e., delta, theta, alpha, beta, and gamma) and VC9.

Procedure

The Aurora Electrode strip with three frontal electrodes was attached to each subject's forehead and connected to the device for brain activity recording. The participants were then asked to complete a standard beginner's task under the surgical simulator Simbionix LAP MENTOR™ (Simbionix, Airport City, Israel), which involved grasping and clamping blood vessels using two different laparoscopic arms. The same task was performed by all subjects. At the end of the task, the participants were rated by the surgical simulator based on three main parameters: accuracy in the execution of the required task, economy of movement (to use as few movements as necessary to produce the technical goal), and time required to accomplish the task. All the behavioral parameters were processed by the virtual simulator Simbionix and were presented in graphs and values at the end of the technical goal. All participants were monitored by the EEG device and were given three attempts to perform the same task. Participants performed three consecutive trials in the same session, with a 5-min break between the trials. The performance of each participant was scored by the surgical simulator's algorithm based on accuracy, economy of movement, and time to complete the task. The concurrent activity of the theta, delta, and VC9 biomarker was extracted by the Aurora system.

Statistical Analysis

Behavioral performance was extracted from the surgical simulator after each trial of each participant, including accuracy (in percentages), economy of movement (in percentages),

TABLE 1 | Coefficients, standard errors, *Z*-scores, *p*-values, and confidence intervals for all effects depicted in the three LMM analyses applied in the three experiments of the present study.

Experiment	Feature	Effect	Coef.	Std. Err.	z	$p > \|z\|$	0.025	0.975
Experiment 1	Economy	Intercept	25.18	1.931	13.039	<0.001	21.395	28.965
		1st trial vs. 2nd trial	6.82	1.892	3.605	<0.001	3.112	10.528
		1st trial vs. 3rd trial	7.968	1.894	4.206	<0.001	4.255	11.681
	Time	Intercept	185.841	5.33	34.867	<0.001	175.395	196.288
		1st trial vs. 2nd trial	−26.789	9.513	−2.816	0.005	−45.434	−8.144
		1st trial vs. 3rd trial	−40.998	8.638	−4.746	<0.001	−57.929	−24.068
	VC9	Intercept	49.531	1.006	49.215	<0.001	47.558	51.503
		1st trial vs. 2nd trial	−1.506	0.594	−2.537	0.011	−2.67	−0.343
		1st trial vs. 3rd trial	−1.75	0.757	−2.313	0.021	−3.233	−0.267
	Theta	Intercept	−13.324	0.625	−21.314	<0.001	−14.549	−12.099
		1st trial vs. 2nd trial	−0.556	0.407	−1.367	0.172	−1.354	0.241
		1st trial vs. 3rd trial	−0.397	0.589	−0.675	0.5	−1.551	0.756
	Delta	Intercept	−2.907	0.742	−3.919	0.000	−4.360	−1.453
		1st trial vs. 2nd trial	−1.493	0.500	−2.989	0.003	−2.472	−0.514
		1st trial vs. 3rd trial	−1.100	0.608	−1.809	0.070	−2.292	0.092
Experiment 2	Economy	Intercept	39.225	2.637	14.873	<0.001	34.056	44.394
		1st trial vs. 2nd trial	−0.525	3.49	−0.15	0.88	−7.365	6.316
		1st trial vs. 3rd trial	4.375	3.412	1.282	0.2	−2.312	11.062
	Time	Intercept	119.596	11.414	10.478	<0.001	97.224	141.968
		1st trial vs. 2nd trial	−8.396	12.449	−0.674	0.5	−32.796	16.004
		1st trial vs. 3rd trial	−22.883	12.971	−1.764	0.078	−48.306	2.54
	VC9	Intercept	49.451	1.341	36.874	<0.001	46.823	52.079
		1st trial vs. 2nd trial	−0.466	0.726	−0.642	0.521	−1.889	0.957
		1st trial vs. 3rd trial	−0.758	0.966	−0.784	0.433	−2.652	1.136
	Theta	Intercept	−13.218	0.824	−16.043	<0.001	−14.833	−11.603
		1st trial vs. 2nd trial	−0.433	0.611	−0.709	0.478	−1.63	0.764
		1st trial vs. 3rd trial	−0.357	0.638	−0.56	0.576	−1.608	0.894
	Delta	Intercept	−2.117	0.953	−2.221	0.026	−3.986	−0.249
		1st trial vs. 2nd trial	−0.123	0.901	−0.137	0.891	−1.888	1.642
		1st trial vs. 3rd trial	−0.862	0.968	−0.891	0.373	−2.759	1.035
Experiment 3	Accuracy	Intercept	59.067	4.008	14.738	<0.001	51.211	66.922
		1st trial vs. 2nd trial	13.023	3.715	3.505	<0.001	5.741	20.305
		1st trial vs. 3rd trial	21.607	3.887	5.559	<0.001	13.99	29.225
	Economy	Intercept	22.003	1.845	11.923	<0.001	18.386	25.62
		1st trial vs. 2nd trial	14.209	1.677	8.47	<0.001	10.921	17.497
		1st trial vs. 3rd trial	17.665	2.305	7.664	<0.001	13.147	22.183
	Time	Intercept	182.022	4.539	40.103	<0.001	173.127	190.918
		1st trial vs. 2nd trial	−31.743	7.613	−4.17	<0.001	−46.664	−16.822
		1st trial vs. 3rd trial	−58.707	5.368	−10.936	<0.001	−69.228	−48.185
	VC9	Intercept	54.489	1.633	33.375	<0.001	51.289	57.689
		1st trial vs. 2nd trial	1.915	1.002	1.911	0.056	−0.049	3.88
		1st trial vs. 3rd trial	2.493	1.967	1.267	0.205	−1.362	6.348
	Theta	Intercept	−6.473	0.899	−7.197	<0.001	−8.235	−4.71
		1st trial vs. 2nd trial	1.285	0.548	2.346	0.019	0.211	2.358
		1st trial vs. 3rd trial	1.432	1.034	1.385	0.166	−0.594	3.459
	Delta	Intercept	5.950	1.438	4.138	0.000	3.132	8.768
		1st trial vs. 2nd trial	2.174	1.157	1.879	0.060	−0.094	4.441
		1st trial vs. 3rd trial	3.609	1.878	1.922	0.055	−0.072	7.289

All LMMs included trial (1/2/3) as a within-participants variable, with encoded effects of Trial 1 vs. Trial 2 and Trial 1 vs. Trial 3.

and time to exceed the trial (in seconds). An initial analysis of all bands revealed that the theta band (averaged dBm at 4–7 Hz), the delta band (averaged dBm at 0.5–4 Hz), and VC9 biomarker (normalized between 1 and 100) were the only three EEG features that were both significant in at least one effect of the LMM analysis and one correlation with the repeated correlation analysis, and therefore were included in the analysis (see **Supplementary Appendix B** for the results of the other oscillations). All features were averaged throughout the task completion. Activity of all the dependent variables was averaged per trial per participant and reported as a mean with a SD. Two main analyses were performed on the behavioral and

EEG data: On each experiment, seven mixed linear models (LMM) (Boisgontier and Cheval, 2016) were designed (one for each dependent variable) to measure the differences between the three simulator trials for each dependent variable. The model used trial numbers (1, 2, or 3) as a within-participants variable, and participants were inserted into the random slope. Since each trial number variable included three levels, indicator variables (aka dummy variables) were computed, with Trial 1 as the reference group. Accordingly, two effects resulted from the model: one from the second trial, which represents the significance of difference between the first and second trials, and one from the third trial, which represents the significance of difference between the first and third trials. These effects were extracted straight from the LMM models, with no need for multiple comparisons corrections. For significant fixed effects, we calculated effect sizes according to Westfall et al. (2014) using an analog of Cohen's d (i.e., the expected mean difference divided by the expected variation of an individual observation). Prior to analyses, each dataset for comparison was revised with Shapiro's test of normal distribution and Levene's test of equal variance (for all test results, see **Supplementary Appendix B**). If any were significant, a Wilcoxon nonparametric test was applied instead of LMM. Two-tailed $p < 0.05$ was considered statistically significant. These analyses were performed using Python Statsmodels (Seabold and Perktold, 2010).

The second analysis was designed to evaluate the correlation between individual performance and the EEG feature. We took into this analysis each of the EEG features' activity and behavioral performance per each participant's trial. Since during each experiment, each participant underwent three task trials, the data points are not independent. To consider both results from repeated participants' trials and between participants, we used the repeated measures correlation (rmcorr, Bakdash and Marusich, 2017). Rmcorr estimates the common regression slope, the association shared among individuals. This analysis was underscored by the rmcorr R package (Bakdash and Marusich, 2017), using 1.4.1717 R studio (RStudio Team, 2021). An initial analysis of all bands revealed that the theta band (averaged dBm at 4–7 Hz), the delta band (averaged dBm at 0.5–4 Hz), and VC9 biomarker (normalized between 1 and 100), were the only three EEG features that were both significant in at least one effect of the LMM analysis and one correlation with the repeated correlation analysis (see **Supplementary Appendix B**). The statistical analyses took into consideration the averaged VC9, delta, and theta activity during each task repetition (an overall average across experiments and trials of 144.17 s).

Results

Behavioral Measurements

A full description of the mixed linear models' (LMM) parameters (coefficients, standard errors, Z-scores, p-values, and confidence intervals) for all models used in this study are presented in **Table 1**. Averages and standard errors of behavioral data in Experiment 1 are presented in **Figure 2**.

A significant increase of the participants' accuracy was observed between the first and second trials and between the first and third trials ($W = 16$, $p = 0.004$; and $W = 10$, $p = 0.004$, respectively), as well as a significant increase in the economy of movement ($p < 0.001$, $d = -0.799$; $p < 0.001$, $d = -1.054$, respectively). The average time required for the completion of the task was also significantly reduced between the first and second trials, and first and third trials ($p = 0.005$, $d = 0.854$; $p < 0.001$, $d = 1.562$, respectively).

Electroencephalography Features

Distributions of VC9, theta, and delta data in Experiment 1 are presented in **Figure 3**. VC9 activity was significantly reduced between the first and second attempts and between the first and third attempts ($p = 0.011$, $d = 0.474$; $p = 0.021$, $d = 0.503$, respectively). Delta activity significantly decreased between the first and second attempts ($p = 0.003$, $d = 0.625$). Theta did not exhibit any significant differences between the trials (all p's > 0.05).

Correlations Between Electroencephalography Features Activity and Behavioral Performance

VC9 activity significantly correlated with all three behavioral measurements and was found to decrease with better performance (i.e., negatively correlated with accuracy and economy of movement and positively correlated with time): $r_{rm} = -0.46$, $p = 0.003$; $r_{rm} = -0.51$, $p = 0.001$; and $r_{rm} = 0.43$, $p = 0.006$ for accuracy, economy of movement, and time, respectively; see **Figure 4**. Delta also significantly correlated with all three measurements: $r_{rm} = -0.57$, $p < 0.001$; $r_{rm} = -0.6$, $p < 0.001$; and $r_{rm} = 0.5$, $p = 0.001$ for accuracy, economy of movement, and time, respectively. Theta decreased with higher accuracy and economy of movement and increased with time: $r_{rm} = -0.41$, $p = 0.01$; $r_{rm} = -0.4$, $p = 0.012$; and $r_{rm} = 0.35$, $p = 0.029$ for accuracy, economy of movement, and time, respectively.

Discussion

Experiment 1 results exhibited a significant improvement in all behavioral measurements of performance, as well as a significant decrease in VC9 and delta activity (but not theta) between the three trials. In addition, VC9 and delta activity decreased with better individual performance, which was expressed by the significant correlations between all three behavioral measures and VC9 and delta activity. Theta also exhibited individual differences and decreased with higher accuracy and economy of movement, and marginally increased with longer time to complete the task. These results suggest that changes in cognitive load correspond to performance in a surgery simulator as shown by VC9, delta, and, to some extent, frontal theta.

Next, we aimed to explore the effect of nighttime sleep memory consolidation on task performance in the simulator and participants' brain activity. Hence, we conducted a second experiment that included 10 participants who had also participated in Experiment 1. They performed an additional three-trial session on the consecutive day.

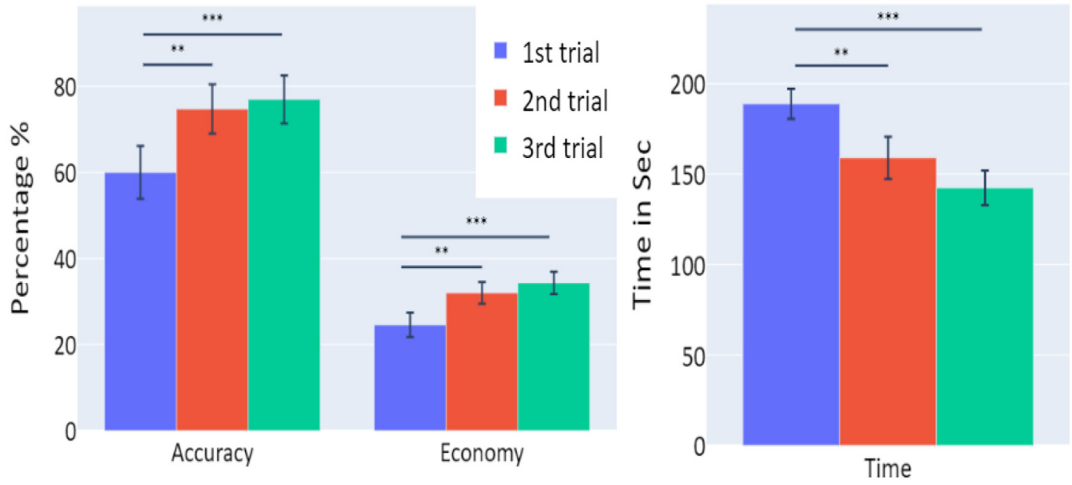


FIGURE 2 | The means of accuracy (in percentage), economy of movement (in percentage), and time to exceed the task (in seconds) obtained in the three trials of Experiment 1, as a function of task repetitions: first trial (blue), second trial (red), and third trial (green), for all participants (n = 19). Error bars represent standard errors. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

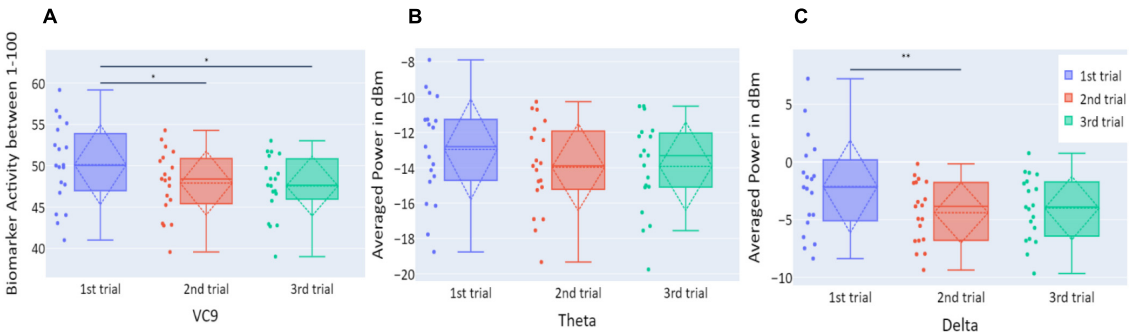


FIGURE 3 | (A) The distribution of participant VC9 activity (normalized between 1 and 100); **(B)** the distribution of theta (averaged power of 4–7 Hz in dBm); and **(C)** the distribution of delta (averaged power of 0.5–4 Hz in dBm), obtained in the three trials of Experiment 1, as a function of task repetitions: first trial (blue), second trial (red), and third trial (green), for all participants (n = 19). Dashed lines represent means and SDs. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

EXPERIMENT 2

Materials and Methods

Participants and Procedure

On the day following Experiment 1, 10 randomly chosen participants performed an additional three trials of the same procedure.

Results

Behavioral Measurements

Behavioral performance did not differ between the three trials of this session in any of the measurements (all p's > 0.05; see **Figure 5**).

Electroencephalography Features

Activity levels of all EEG features (VC9, theta, and delta) did not differ between the trials of the session ($p > 0.05$; see **Figure 6**).

Correlations Between Electroencephalography Features Activity and Behavioral Performance

VC9 activity significantly correlated with all three behavioral measurements and was found to decrease with better participant performance (i.e., negatively correlated with accuracy and economy of movement and positively correlated with time): $r_{rm} = -0.47$, $p = 0.033$; $r_{rm} = -0.67$, $p = 0.001$; and $r_{rm} = 0.64$, $p = 0.002$ for accuracy, economy of movement, and time, respectively. Delta also significantly correlated with all three measurements: $r_{rm} = -0.69$, $p = 0.001$; $r_{rm} = -0.65$, $p = 0.001$; and $r_{rm} = 0.63$, $p = 0.002$ for accuracy, economy of movement, and time, respectively. Theta did not correlate significantly with any of the behavioral measurements (all p's > 0.05; see **Figure 7**).

Comparison Between Experiment 1 and Experiment 2

To explore the "offline gains" between Experiment 1 and Experiment 2, behavioral performance and brain activity of the 10 participants who participated in both experiments were studied. Paired t-tests on the activity of the 10 participants

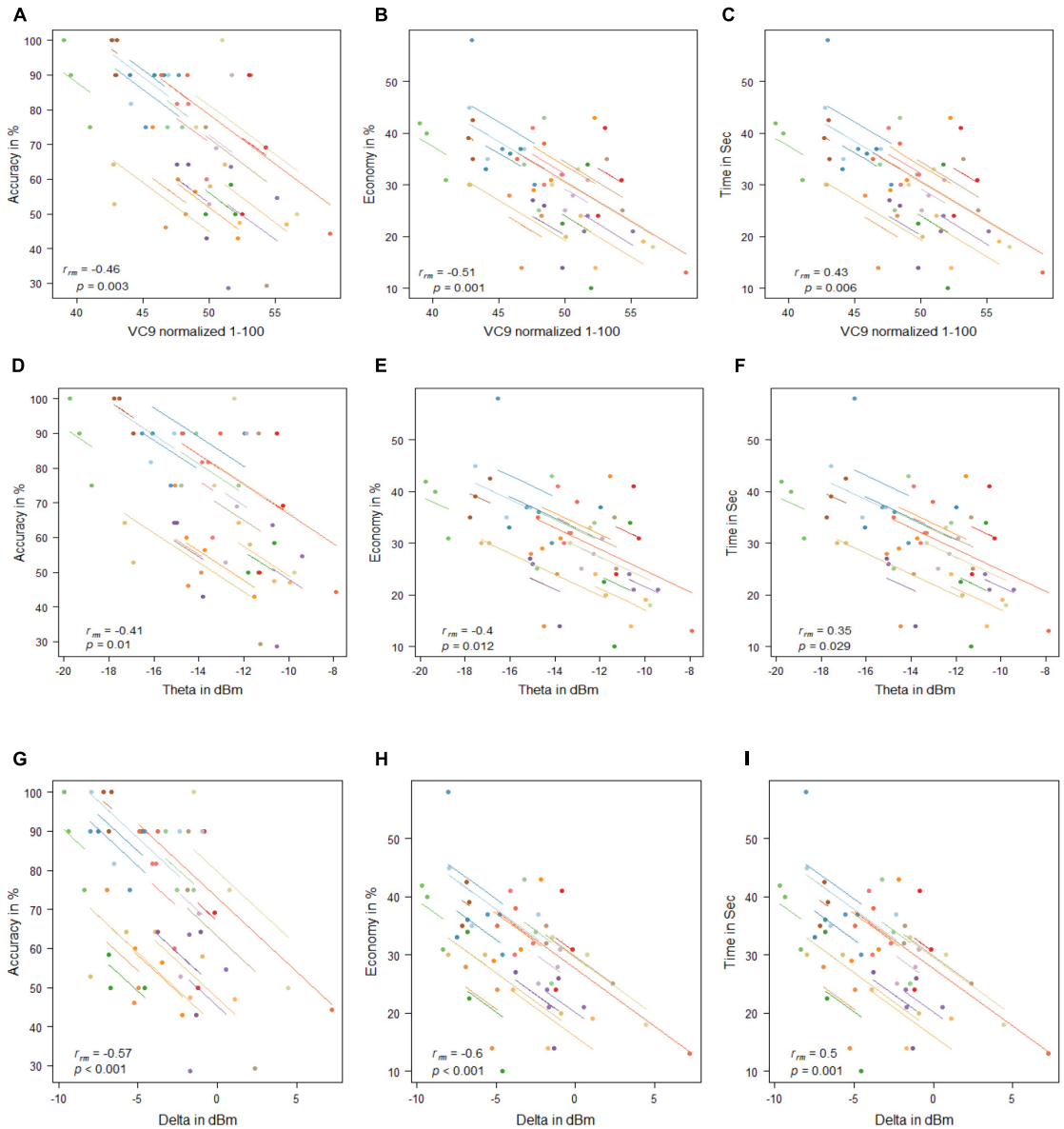


FIGURE 4 | Mean activity of VC9 as a function of accuracy **(A)**, economy **(B)**, and time **(C)**; theta as a function of accuracy **(D)**, economy **(E)**, and time **(F)**; and delta as a function of accuracy **(G)**, economy **(H)**, and time **(I)** per trial and participant obtained in Experiment 1 (n = 19). R_{rm} and p-values presented.

who underwent both experiments were calculated. The t-tests compared accuracy, economy of movement, time, VC9, theta, and delta activity for the last trial of Experiment 1 and the first trial of Experiment 2. Both behavioral performance and brain activity were the same in the first trial of Experiment 2 relative to the last trial of Experiment 1: all p's > 0.05; see **Figure 8** and **Table 2**.

Discussion

The results of Experiment 2 exhibited a clear correspondence between behavioral performance and EEG features. First, both behavioral performance and all EEG features' activity did not exhibit significant differences between the three trials, meaning that participants' performance did not improve, and theta, delta, and VC9 activity did not decrease between the three trials. Second, VC9 and delta activity decreased with better behavioral performance, as depicted by a significant correlation with all three behavioral measurements. Theta correlated with accuracy but not with economy of movement and time.

The comparison between the last trial of Experiment 1 and the first trial of Experiment 2 showed no difference between the two in any of the behavioral or neurophysiological measures. However, taken together with the null improvement during these three sessions, this lack of difference may be explained by participants having reached their asymptotic level (ceiling effect): that participants reached their maximum ability on the third trial

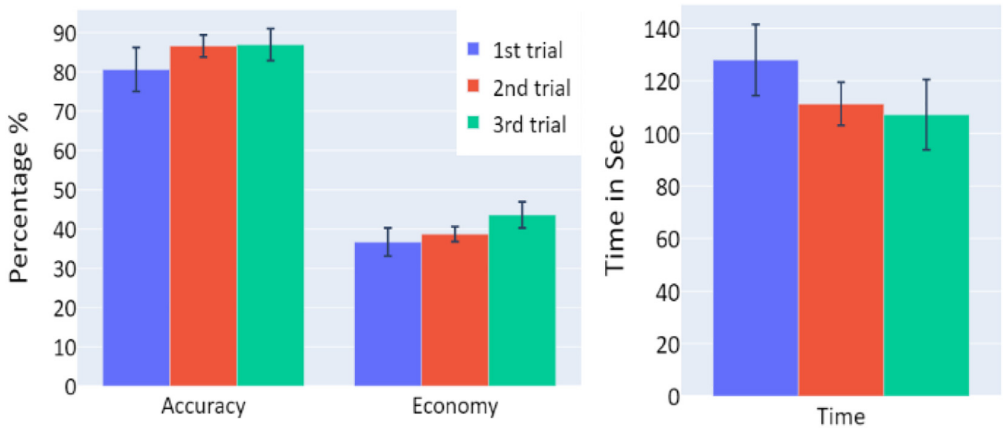


FIGURE 5 | The means of accuracy (in percentage), economy of movement (in percentage), and time to exceed the task (in seconds) obtained in the three trials of Experiment 2, as a function of task repetitions: first trial (blue), second trial (red), and third trial (green), for participants who underwent Experiment 2 (n = 10). Error bars represent standard errors. *p < 0.05, **p < 0.01, and ***p < 0.001.

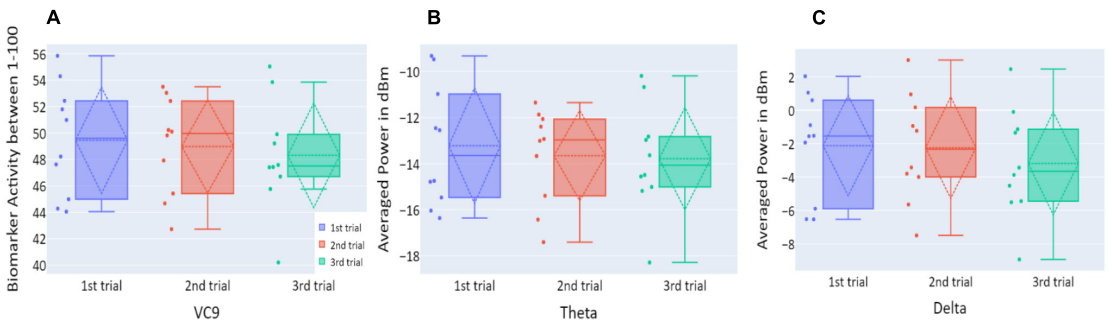


FIGURE 6 | (A) The distribution of user VC9 activity (normalized between 1 and 100); **(B)** the distribution of theta (averaged power of 4–7 Hz in dBm); and **(C)** the distribution of delta (averaged power of 0.5–4 Hz in dBm), obtained in the three trials of Experiment 2, as a function of task repetitions: first trial (blue), second trial (red), and third trial (green), for all participants (n = 10). Dashed lines represent means and SDs. *p < 0.05, **p < 0.01, and ***p < 0.001.

of the first day, and maintained it through the first trials of the second day. To further reveal "offline gains" from nighttime sleep consolidation, we conducted a third experiment on an additional 19 medical students who had not participated in Experiments 1 and 2. Participants performed a single-task trial per day for three consecutive days. This was done to make sure participants did not reach asymptotic performance levels before the nighttime sleep consolidation, since they would only complete a single trial before nighttime sleep.

EXPERIMENT 3

Materials and Methods

Participants and Procedure

Nineteen (63% females) healthy medical students from their first to sixth years of studying (mean = 4), who did not participate in former Experiments 1 and 2, with mean age of 25.631 (SD = 2.532), participated in Experiment 3. All students had no prior experience using a surgical simulator. Participants underwent the same task as in Experiments 1 and 2, with one trial per day over three consecutive days.

Results

Behavioral Measurements

Averages and standard errors of the behavioral data in Experiment 3 are presented in **Figure 9**. A significant increase of the participants' accuracy was observed between the first and second trials and between the first and third trials (W = 11, p = 0.002; W = 2, p < 0.001, respectively), as well as a significant increase in the economy of movement (p < 0.001, d = −1.762; p < 0.001, d = −1.83, respectively). The average time required for the completion of the task was also significantly reduced between the trials (p < 0.001, d = 1.014; p < 0.001, d = 3.171 for first vs. second trials and first vs. third trials, respectively).

Electroencephalography Features

Distributions of VC9, theta, and delta data in Experiment 3 are presented in **Figure 10**. There were no significant differences in VC9 and delta activity between the trials on consecutive days (all

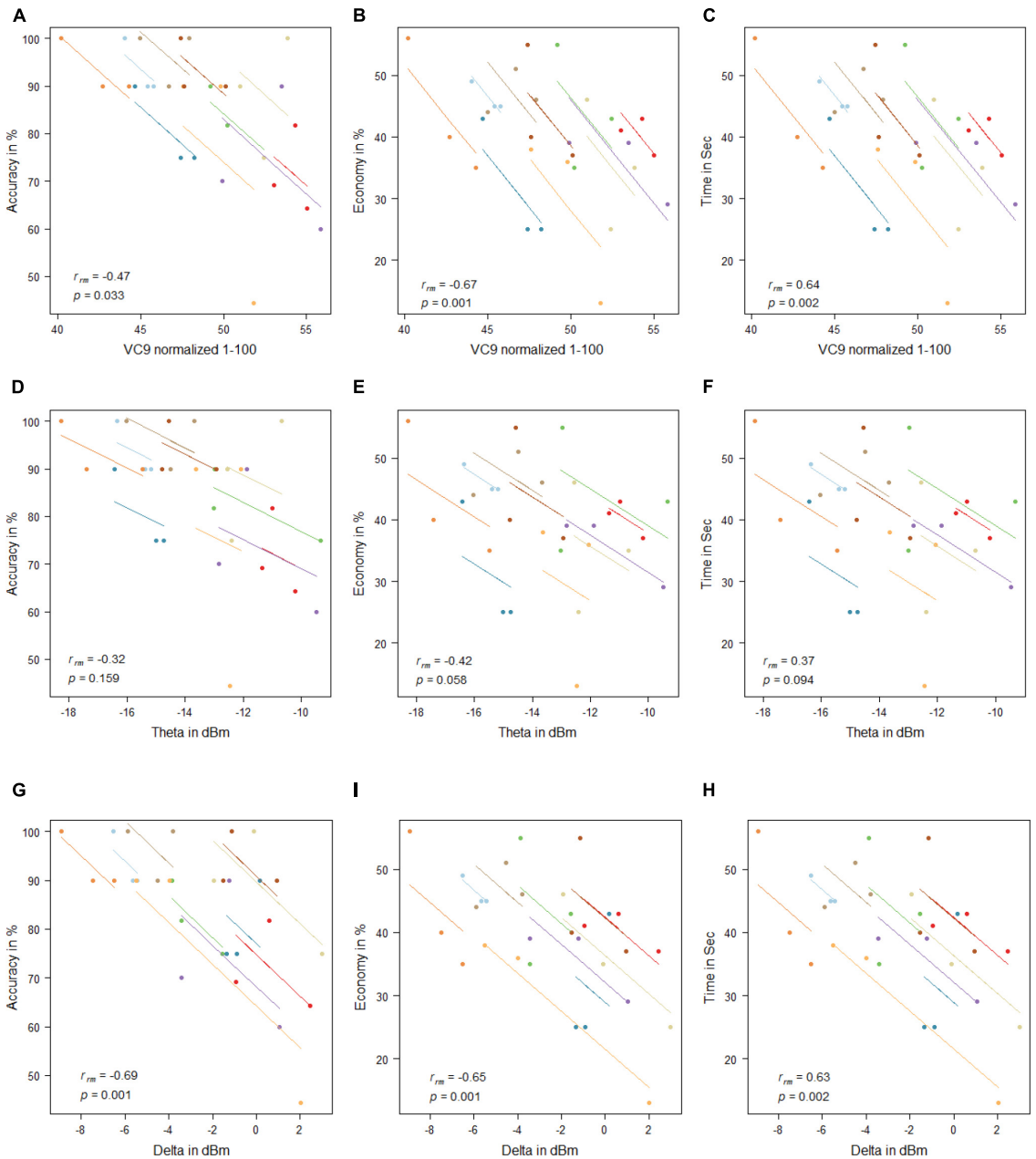


FIGURE 7 | Mean activity of VC9 as function of accuracy **(A)**, economy **(B)**, and time **(C)**; theta as function of accuracy **(D)**, economy **(E)**, and time **(F)**; and delta as function of accuracy **(G)**, economy **(H)**, and time **(I)** per trial and participant obtained in Experiment 2 (n = 10). R_{rm} and p-values presented.

p's > 0.05). Theta exhibited a significant difference between the first and second trials (p = 0.019, d = −0.354), but the difference between the first and third trials did not reach significance level (p > 0.05).

Correlations Between Electroencephalography Features Activity and Behavioral Performance

No significant correlations were found between any of the EEG features and behavioral measurement (all p's > 0.05; see **Figure 11**).

Discussion

Experiment 3 results reflect "offline gains": participants' performances under the simulator improved with the trials' repetition, although each occurred on the consecutive day. However, VC9, delta, and theta did not show significant differences between the trials, nor did they exhibit significant correlations with any of the behavioral measurements. This discrepancy might be explained by the different brain networks that take part in "online" versus "offline" learning; see general discussion below.

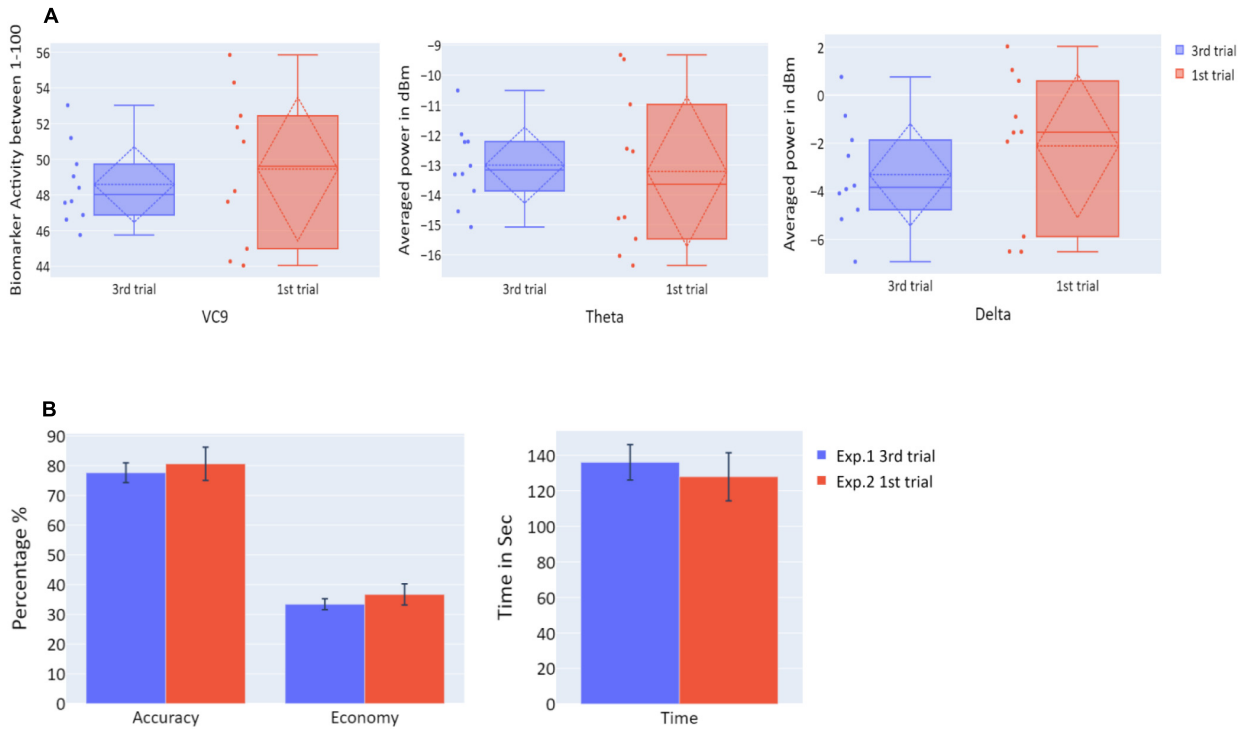


FIGURE 8 | (A) The distribution of VC9 (normalized between 1 and 100), theta (averaged power of 4–7 Hz in dBm), and delta (averaged power of 0.5–4 Hz in dBm), obtained in the last trial repetition of Experiment 1 (blue), and the first task repetition of Experiment 2 (red), for participants who participated in both Experiment 1 and Experiment 2 (n = 10). Dashed lines represent means and SDs. **(B)** The means of accuracy (in percentage), economy of movement (in percentage), and time to exceed the task (in seconds), obtained in the last trial repetition of Experiment 1 (blue), and the first task repetition of Experiment 2 (red), for participants who participated in both Experiment 1 and Experiment 2 (n = 10). Error bars represent standard errors.

GENERAL DISCUSSION

In this study, we continuously measured cognitive load levels using theta and band power and VC9 biomarker activity during task performance on a surgical simulator. A single-channel EEG device utilizing decomposition of the EEG signal *via* harmonic analysis was used to obtain the novel VC9 biomarker as well as the frequently used theta and delta bands. All EEG features (i.e., theta, delta, and VC9) were previously shown to correlate with cognitive load levels (Antonenko et al., 2010; Maimon et al., 2020; Bolton et al., 2021; Molcho et al., 2021). Experiment 1 consisted of three repeating trials on the surgery simulator, and its results showed that participants' performances improved while VC9 and delta activity decreased. Additionally, VC9, delta, and, to some extent, theta band decreased with higher individual performance. Overall, these results support previous findings that as the participant becomes more proficient in performing a task, as depicted by the behavioral performance, the pre-frontal activity associated with cognitive load is reduced (Takeuchi et al., 2013).

To examine the effect of nighttime sleep memory consolidation, we conducted Experiment 2, which included three simulator trials on the consecutive day. Testing the "offline gains" refers to the improvements in skill acquisition preceding a consolidation during nighttime sleep between tasks trainings and without further practice (Walker et al., 2002). Results revealed no significant differences in performance or EEG features between the three trials. Although they did not exhibit significant decreases between the trials, the VC9, delta, and, to some extent, theta still exhibited significant correlations with individual performances. Further analysis revealed that both behavioral and EEG features results maintained the same levels between the last trial of the first day and the first trial on the second day. Taken together, these findings potentially indicate that "offline gains" were not detected, since participants reached their maximum performance levels on the last trial of the first day and maintained them throughout the trials on the following day.

Consequently, we conducted an additional experiment in which 19 medical students performed a single simulator trial on three consecutive days. This was done to prevent participants from reaching their maximum performance levels on the first

TABLE 2 | t/W statistics and p-values for t-tests/Wilcoxon tests comparing Experiment 1 last trial and Experiment 2 first trial, for participants who participated in both Experiment 1 and Experiment 2 (n = 10).

Feature	Statistic	p-value
Accuracy	$t = -0.476$	0.645
Economy	$t = -0.831$	0.428
Time	$t = 0.524$	0.613
VC9	$W = 21$	0.557
Theta	$W = 26$	0.922
Delta	$t = 1.103$	0.298

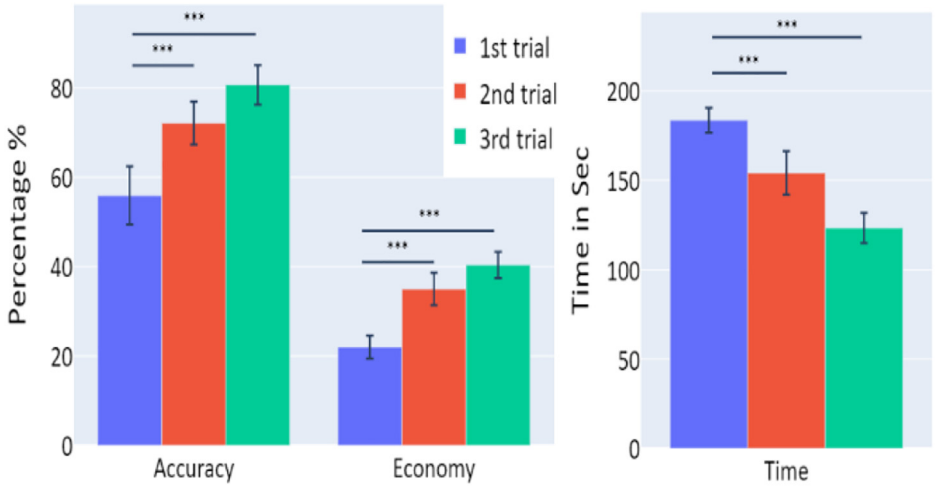


FIGURE 9 | The means of accuracy (in percentage), economy of movement (in percentage), and time to exceed the task (in seconds) obtained in the three trials of Experiment 3, as a function of task repetitions: first trial (blue), second trial (red), and third trial (green), for participants who underwent Experiment 3 (n = 19). Error bars represent standard errors. *p < 0.05, **p < 0.01, and ***p < 0.001.

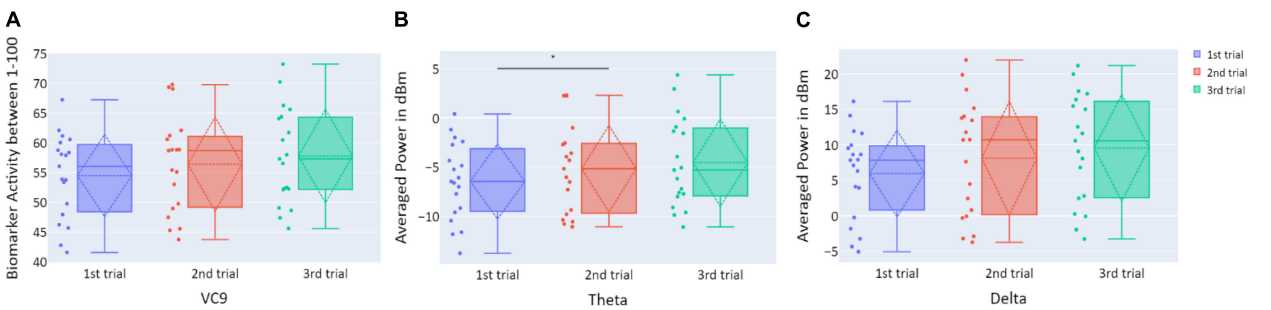


FIGURE 10 | (A) The distribution of user VC9 activity (normalized between 1 and 100); **(B)** the distribution of theta (averaged power of 4–7 Hz in dBm); and **(C)** the distribution of delta (averaged power of 0.5–4 Hz in dBm), obtained in the three trials of Experiment 3, as a function of task repetitions: first trial (blue), second trial (red), and third trial (green), for all participants (n = 19). Dashed lines represent means and SDs. *p < 0.05, **p < 0.01, and ***p < 0.001.

day, and to reveal the nighttime sleep memory consolidation related to "offline gains." Indeed, participants' performances in all behavioral measurements improved significantly between the testing days. Interestingly, VC9 and delta showed no difference between the trials, and the theta band even showed a mild increase in activity between the trials. This lack of increase in activity can be explained by the difference in brain networks that are involved in fast vs. slow stages of motor skill learning (Dayan and Cohen, 2011). Similar to online learning, fast motor skill acquisition occurs during task training and could last minutes (Karni et al., 1995). In the slow stage, further gains are achieved across multiple sessions of training, mostly divided by nighttime sleep consolidation. The neural substrates during the fast stage show complex brain activation patterns. First, they include increasing activity in the **Supplementary Motor Area** (SMA), dorsomedial striatum (DMS), premotor cortex (PM), posterior parietal cortex (PPC), and posterior cerebellum (Honda et al., 1998; Grafton et al., 2002; Floyer-Lea and Matthews, 2005). This reflects the requirement of additional cortical brain activity during practice. At the same time, the fast stage also causes decreased activity of the dorsolateral prefrontal cortex (DLPFC), primary motor cortex (M1), and pre **Supplementary Motor Area** (preSMA, Poldrack, 2000). These decreases may suggest that with online practice, one uses fewer neuronal resources (Hikosaka et al., 2002). Conversely, the slow motor skill stage is characterized by increased activity in M1, S1, SMA, and DLS (Floyer-Lea and Matthews, 2005), and decreased activity in the lateral cerebellum (Lehéricy et al., 2005). Notably, the decrease in the frontal area (DLPFC), which commonly correlates with WM load (Manoach et al., 1997), is not visible during this slow learning stage. Thus, progress from fast to slow stages can be generalized to the shift from anterior to more posterior brain regions (Floyer-Lea and Matthews, 2005).

The current WM biomarkers that were used in the present research were extracted *via* frontal single-EEG channel, and although this electrode can extract additional brain activity beyond the frontal lobe, we aimed to focus on biomarkers

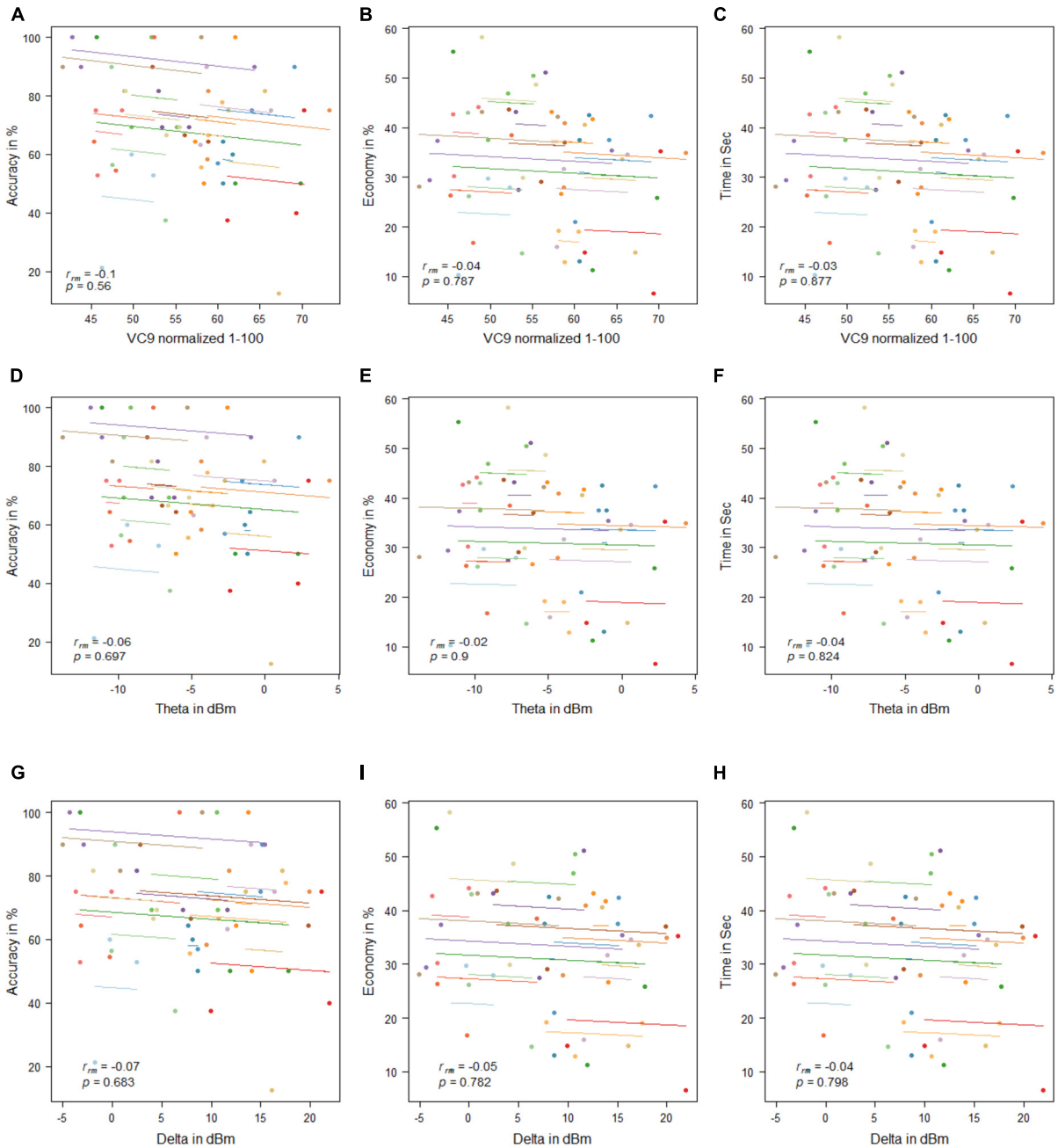


FIGURE 11 | Mean activity of VC9 as a function of accuracy **(A)**, economy **(B)**, and time **(C)**; theta as a function of accuracy **(D)**, economy **(E)**, and time **(F)**; and delta as a function of accuracy **(G)**, economy **(H)**, and time **(I)** per trial and participant obtained in Experiment 3 (n = 19). R_{rm} and p-values presented.

that reflect frontal activity associated with WM load. Taken together, this may explain the results obtained in Experiment 3 of the present study, which revealed prominent "offline gains" within the behavioral performance of the task, but no decrease within the current EEG WM load biomarkers. Further research, however, should look for novel biomarkers that will be able to correlate with such posterior/limbic activity. Finally, this discrepancy may further support frontal theta, delta, and VC9 as continuous measurements of cognitive load, to monitor WM load during task completion and not as evaluators of cognitive load between tasks' sessions. Therefore, these biomarkers may be adequate for assessing laparoscopic dexterity of non-experts during their first real-life surgeries. However, for the assessment of surgeons' improvement in between laparoscopic practices, other biomarkers should be considered.

The activation patterns obtained in Experiments 1 and 2 are compatible with WM load, as previously described

(Zarjam et al., 2011; Wang et al., 2016). Delta and VC9 activity decreased with the lower cognitive load that resulted from simulator trial repetition, but the theta band activity did not decrease. As shown in previous studies (Maimon et al., 2020; Molcho et al., 2021), VC9 was found to be more sensitive to lower loads than theta and delta, as its activity also showed a significant decrease between the first and third task trials. This further validates VC9 as an effective biological measurement for the assessment of cognitive load while performing laparoscopic tasks using the surgical simulator specifically and suggests that it may provide a measure of cognitive functioning during surgery. Due to the unobtrusive nature of mobile EEG devices, such devices can be used by surgeons during live operations. Intraoperative evaluation can provide an objective metric to assess surgeons' performance in real-life scenarios and measure additional unobservable behaviors, such as mental readiness (Cha and Yu, 2021).

Since simulation-based training is an effective tool for acquiring practical skills, the question remains as to which methods should be utilized to optimize this process and achieve better assessment of improved manual dexterity. Specifically, it is not fully understood which brain processes during simulation-based training translate to better skill acquisition through practice. Here, we show that VC9 and delta, collected *via* single-channel EEG set, can be reliably used to monitor and assess participants' WM levels during manual practice. Furthermore, the correlations between VC9, delta, and improved simulator scores can be translated to other areas and used in other procedures that require manual training. Importantly, as this device is portable and can be easily worn inside the operating room, it could potentially be used to predict participants' WM loads and indirect performance during real-life operating procedures.

This study has several limitations. The analyses were performed on young medical internists, which may not accurately reflect the overall medical population. Additionally, evaluating finer differences between the experiments requires a larger cohort of participants. Future studies on a larger and more diverse population should further validate the findings presented in this work, as well as study the effect of prolonged breaks on the activity of these biomarkers. Moreover, a 24-h break period between the sessions may not be sufficient to study the effects of long-term memory on the prefrontal brain activity during the simulation. Future studies should evaluate longer time windows to assess the temporal changes in the activity of the working-memory-associated biomarkers and search for additional biomarkers to represent other cerebral regions beyond the frontal regions (like the limbic or motor systems).

To conclude, in this study, we extracted the previously validated cognitive load biomarkers—theta, delta, and VC9—from a novel single-channel EEG using advanced signal analysis. Results showed high correlations between the EEG features and participants' individual performances using a surgical simulator. As surgical simulations allow doctors to gain important skills and experience needed to perform procedures without any patient risk, evaluation and optimization of these effects on medical staff are crucial. This could potentially be expanded to evaluate the efficacy of different medical simulations to provide more efficient training to medical staff and to measure cognitive and mental loads in real laparoscopic surgeries.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data that support the findings of this study are available for simbionix and Neurosteer Ltd. Restrictions apply to the availability of these data, which were used under license for this study. Requests to access the datasets should be directed to lior@neurosteer.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Galilee Medical Center Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AB, MB, NM, and NI conceived and planned the experiments. AB, MB, and DD carried out the experiments. NI and LM designed and provided tech support for the EEG device. MB, NM, and LM contributed to the interpretation of the results. MB and NM took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.694010/full#supplementary-material

REFERENCES

Andersen, S. A. W., Konge, L., and Sørensen, M. S. (2018). The effect of distributed virtual reality simulation training on cognitive load during subsequent dissection training. *Med. Teacher* 40, 684–689.

Andersen, S. A. W., Mikkelsen, P. T., Konge, L., Cayé-Thomasen, P., and Sørensen, M. S. (2016). The effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills: a randomized controlled trial. *Adv. Simul.* 1, 1–8.

Antonenko, P., Paas, F., Grabner, R., and van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educ. Psychol. Rev.* 22, 425–438. doi: 10.1007/s10648-010-9130-y

Bakdash, J. Z., and Marusich, L. R. (2017). Repeated measures correlation. *Front. Psychol.* 8:456. doi: 10.3389/fpsyg.2017.00456

Bastiaansen, M. C., Posthuma, D., Groot, P. F., and De Geus, E. J. (2002). Event-related alpha and theta responses in a visuo-spatial working memory task. *Clin. Neurophysiol.* 113, 1882–1893. doi: 10.1016/S1388-2457(02)00303-6

Boisgontier, M. P., and Cheval, B. (2016). The ANOVA to mixed model transition. *Neurosci. Biobehav. Rev.* 68, 1004–1005. doi: 10.1016/j.neubiorev.2016.05.034

Bolton, F., Te'Eni, D., Maimon, N. B., and Toch, E. (2021). "Detecting interruption events using EEG," in *Proceedings of the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)* (Nara: IEEE), 33–34.

Bong, C. L., Lightdale, J. R., Fredette, M. E., and Weinstock, P. (2010). Effects of simulation versus traditional tutorial-based training on physiologic stress levels among clinicians: a pilot study. *Simul. Healthcare* 5, 272–278. doi: 10.1097/sih.0b013e3181e98b29

Cha, J. S., and Yu, D. (2021). Objective measures of surgeon nontechnical skills in surgery: a scoping review. *Hum. Fact.* doi: 10.1177/0018720821995319

Coifman, R. R., and Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theory* 38, 713–718. doi: 10.1109/18.119732

Constantinidis, C., and Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nat. Rev. Neurosci.* 17:438. doi: 10.1038/nrn.2016.43

Dayan, E., and Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron* 72, 443–454. doi: 10.1016/j.neuron.2011.10.008

Di Flumeri, G., Borghini, G., Aricò, P., Sciaraffa, N., Lanzi, P., and Pozzi, S. (2018). EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Front. Hum. Neurosci.* 12:509. doi: 10.3389/fnhum.2018.00509

Dussault, C., Jouanin, J. C., Philippe, M., and Guezennec, C. Y. (2005). EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviation Space Environ. Med.* 76, 344–351.

Erdfelder, E., Faul, F., and Buchner, A. (1996). GPOWER: a general power analysis program. Behavior research methods, *Instruments Comp.* 28, 1–11.

Floyer-Lea, A., and Matthews, P. M. (2005). Distinguishable brain activation networks for short- and long-term motor skill learning. *J. Neurophysiol.* 94, 512–518. doi: 10.1152/jn.00717.2004

Fraser, K. L., Ayres, P., and Sweller, J. (2015). Cognitive load theory for the design of medical simulations. *Simul. Healthcare* 10, 295–307. doi: 10.1097/sih.0000000000000097

Frederiksen, J.G., Sørensen, S. M. D., Konge, L. Svendsen, M. B. S., Nobel-Jørgensen M., Bjerrum F. et al. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surg. Endosc.* 34, 1244–1252. doi: 10.1007/s00464-019-06887-8

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., and Moses, G. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann. Surg.* 241:364. doi: 10.1097/01.sla.0000151982.85062.80

Gauger, P. G., Hauge, L. S., Andreatta, P. B., Hamstra, S. J., Hillard, M. L., and Arble, E. P., (2010). Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. *Am. J. Surg.* 199, 72–80. doi: 10.1016/j.amjsurg.2009.07.034

Gevins, A., Smith, M. E., McEvoy, L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb. Cortex* 7, 374–385. doi: 10.1093/cercor/7.4.374

Grafton, S. T., Hazeltine, E., and Ivry, R. B. (2002). Motor sequence learning with the nondominant left hand. *Exp. Brain Res.* 146, 369–378. doi: 10.1007/s00221-002-1181-y

Gurusamy, K. S., Aggarwal, R., Palanivelu, L., and Davidson, B. R. (2009). Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Datab. Syst. Rev.* 1:CD006575.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Front. Integrat. Neurosci.* 7:83. doi: 10.3389/fnint.2013.00083

Harmony, T., Fernández, T., Silva, J., Bernal, J., Díaz-Comas, L., Reyes, A., et al. (1996). EEG delta activity: an indicator of attention to internal processing during performance of mental tasks. *Int. J. Psychophysiol.* 24, 161–171. doi: 10.1016/S0167-8760(96)00053-0

Hikosaka, O., Rand, M. K., Nakamura, K., Miyachi, S., Kitaguchi, K., Sakai, K., (2002). Long-term retention of motor skill in macaque monkeys and humans. *Exp. Brain Res.* 147, 494–504. doi: 10.1007/s00221-002-1258-7

Honda, M., Deiber, M. P., Ibánez, V., Pascual-Leone, A., Zhuang, P., and Hallett, M. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning. A PET study. *Brain J. Neurol.* 121, 2159–2173.

Issenberg, S. B., Ringsted, C., Østergaard, D., and Dieckmann, P. (2011). Setting a research agenda for simulation-based healthcare education: a synthesis of the outcome from an Utstein style meeting. *Simul. Healthcare* 6, 155–167. doi: 10.1097/sih.0b013e3182207c24

Jolles, D. D., van Buchem, M. A., Crone, E. A., and Rombouts, S. A. (2013). Functional brain connectivity at rest changes after working memory training. *Hum. Brain Mapp.* 34, 396–406. doi: 10.1002/hbm.21444

Karni, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., and Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature* 377, 155–158. doi: 10.1038/377155a0

Kim, J. H., Kim, D. W., and Im, C. H. (2017). Brain areas responsible for vigilance: an EEG source imaging study. *Brain Topograp.* 30, 343–351. doi: 10.1007/s10548-016-0540-0

Kunkler, K. (2006). The role of medical simulation: an overview. *Int. J. Med. Robot. Comp. Assisted Surg.* 2, 203–210. doi: 10.1002/rcs.101

Larsen, C. R., Soerensen, J. L., Grantcharov, T. P., Dalsgaard, T., Schouenborg, L., Ottosen, C., and Ottesen, B. S. (2009). Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *BMJ* 338:b2074. doi: 10.1136/bmj.b2074

Lehéricy, S., Benali, H., Van de Moortele, P. F., Pélégrini-Issac, M., Waechter, T., Ugurbil, K., and Doyon, J. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12566–12571. doi: 10.1073/pnas.0502762102

Lugassy, D., Herszage, J., Pilo, R., Brosh, T., and Censor, N. (2018). Consolidation of complex motor skill learning: evidence for a delayed offline process. *Sleep* 41:zsy123. doi: 10.1093/sleep/zsy123

Maimon, N. B., Molcho, L., Intrator, N., and Lamy, D. (2020). Single-channel EEG features during n-back task correlate with working memory load. arXiv [Preprint]. arXiv:2008.04987.

Maimon, N. B., Molcho, L., Intrator, N., and Lamy, D. (2021). "Novel single-channel EEG features correlate with working memory load," in *Proceedings of the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, Nara, 471–472. doi: 10.1109/LifeTech52111.2021.9391963

Manoach, D. S., Schlaug, G., Siewert, B., Darby, D. G., Bly, B. M., Benfield, A., Edelman, R. R., and Warach, S. (1997). Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport* 8, 545–549. doi: 10.1097/00001756-199701200-00033

Miller, J. G. (1956). General behavior systems theory and summary. *J. Counsel. Psychol.* 3:120.

Molcho, L., Maimon, N. B., Pressburger, N., Regev-Plotnik, N., Rabinowicz, S., Intrator, N., and Sasson, A. (2021). Detection of cognitive decline using a single-channel EEG system with an interactive assessment tool. *Res. Square [Preprint]* doi: 10.21203/rs.3.rs-242345/v1

Neretti, N., and Intrator, N. (2002). "An adaptive approach to wavelet filters design," in *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, Martigny, 317–326.

Paas, F., Tuovinen, J. E., Tabbers, H., and Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* 38, 63–71. doi: 10.1207/S15326985EP3801_8

Poldrack, R. A. (2000). Imaging brain plasticity: conceptual and methodological issues—a theoretical review. *Neuroimage*, 12, 1–13. doi: 10.1006/nimg.2000.0596

RStudio Team (2021). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, PBC.

Schapkin, S. A., Raggatz, J., Hillmert, M., and Böckelmann, I. (2020). EEG correlates of cognitive load in a multiple choice reaction task. *Acta Neurobiol. Exp.* 80, 76–89.

Scharinger, C., Soutschek, A., Schubert, T., and Gerjets, P. (2017). Comparison of the working memory load in n-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. *Front. Hum. Neurosci.* 11:6.

Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., and Bastiaansen, M. C. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage*, 44, 1224–1238. doi: 10.1016/j.neuroimage.2008.08.041

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84:1.

Seabold, S., and Perktold, J. (2010). "Statsmodels: econometric and statistical modeling with python", in *Proceedings of the 9th Python in Science Conference*, Vol. 57, Austin, TX, 61

Takeuchi, H., Taki, Y., Nouchi, R., Hashizume, H., Sekiguchi, A., Kotozaki, Y., and Kawashima, R. (2013). Effects of working memory training on functional connectivity and cerebral blood flow during rest. *Cortex* 49, 2106–2125. doi: 10.1016/j.cortex.2012.09.007

Thompson, J. E., Egger, S., Böhm, M., Haynes, A. M., Matthews, J., Rasiah, K., and Stricker, P. D. (2014). Superior quality of life and improved surgical margins are achievable with robotic radical prostatectomy after a long learning curve: a prospective single-surgeon study of 1552 consecutive cases. *Eur. Urol.* 65, 521–531. doi: 10.1016/j.eururo.2013.10.030

Tucker, M. A., Morris, C. J., Morgan, A., Yang, J., Myers, S., Pierce, J. G., and Scheer, F. A. (2017). The relative impact of sleep and circadian drive on motor skill acquisition and memory consolidation. *Sleep* 40:zsx036. doi: 10.1093/sleep/zsx036

Van Dillen, L. F., Heslenfeld, D. J., and Koole, S. L. (2009). Tuning down the emotional brain: an fMRI study of the effects of cognitive load on the processing of affective images. *Neuroimage* 45, 1212–1219. doi: 10.1016/j.neuroimage.2009.01.016

Van Gerven, P. W., Paas, F., Van Merriënboer, J. J., and Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiol.* 41, 167–174. doi: 10.1111/j.1469-8986.2003.00148.x

Van Gog, T., and Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learn. Instruction* 20, 95–99. doi: 10.1016/j.learninstruc.2009.02.009

Van Merriënboer, J. J., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177. doi: 10.1007/s10648-005-3951-0

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., and Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* 35, 205–211. doi: 10.1016/s0896-6273(02)00746-8

Wang, S., Gwizdka, J., and Chaovalitwongse, W. A. (2016). Using wireless EEG signals to assess memory workload in the n-back task. *IEEE Trans. Hum. Machine Syst.* 46, 424–435. doi: 10.1109/THMS.2015.2476818

Westfall, J., Kenny, D. A., and Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exper. Psychol. Gen.* 143:2020. doi: 10.1037/xge0000014

Ye, J., and Li, Q. (2005). A two-stage linear discriminant analysis via QR-decomposition. *IEEE Trans. Pattern Anal. Machine Intell.* 27, 929–941. doi: 10.1109/TPAMI.2005.110

Yiannakopoulou, E., Nikiteas, N., Perrea, D., and Tsigris, C. (2015). Virtual reality simulators and training in laparoscopic surgery. *Int. J. Surg.* 13, 60–64. doi: 10.1016/j.ijsu.2014.11.014

Zarjam, P., Epps, J., and Chen, F. (2011). "Characterizing working memory load using EEG delta activity," in *Proceedings of the 2011 19th European Signal Processing Conference*, Barcelona, 1554–1558

Conflict of Interest: NM is a part-time researcher at Neurosteer Ltd., which supplied the EEG system. LM is a clinical researcher at Neurosteer Ltd. NI is the founder of Neurosteer 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.

Copyright © 2022 Maimon, Bez, Drobot, Molcho, Intrator, Kakiashvilli and Bickel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 17 February 2022
doi: 10.3389/fnins.2022.756938

A Novel Constrained Non-negative Matrix Factorization Method for Group Functional Magnetic Resonance Imaging Data Analysis of Adult Attention-Deficit/Hyperactivity Disorder

Ying Li[1], Weiming Zeng[1], Yuhu Shi[1], Jin Deng[2], Weifang Nie[1], Sizhe Luo[1] and Jiajun Yang[3]*

[1] Lab of Digital Image and Intelligent Computation, Shanghai Maritime University, Shanghai, China, [2] College of Mathematics and Information, South China Agricultural University, Guangzhou, China, [3] Department of Neurology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China

OPEN ACCESS

Edited by:
Kaijian Xia,
Changshu No. 1 People's Hospital, China

Reviewed by:
Linling Li,
Shenzhen University, China
Shanbao Tong,
Shanghai Jiao Tong University, China

***Correspondence:**
Weiming Zeng
zengwm86@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *11 August 2021*
Accepted: *10 January 2022*
Published: *17 February 2022*

Citation:
Li Y, Zeng W, Shi Y, Deng J, Nie W, Luo S and Yang J (2022) A Novel Constrained Non-negative Matrix Factorization Method for Group Functional Magnetic Resonance Imaging Data Analysis of Adult Attention-Deficit/Hyperactivity Disorder. Front. Neurosci. 16:756938. doi: 10.3389/fnins.2022.756938

Attention-deficit/hyperactivity disorder (ADHD) is a common childhood psychiatric disorder that often persists into adulthood. Extracting brain networks from functional magnetic resonance imaging (fMRI) data can help explore neurocognitive disorders in adult ADHD. However, there is still a lack of effective methods to extract large-scale brain networks to identify disease-related brain network changes. Hence, this study proposed a spatial constrained non-negative matrix factorization (SCNMF) method based on the fMRI real reference signal. First, non-negative matrix factorization analysis was carried out on each subject to select the brain network components of interest. Subsequently, the available spatial prior information was mined by integrating the interested components of all subjects. This prior constraint was then incorporated into the NMF objective function to improve its efficiency. For the sake of verifying the effectiveness and feasibility of the proposed method, we quantitatively compared the SCNMF method with other classical algorithms and applied it to the dynamic functional connectivity analysis framework. The algorithm successfully extracted ten resting-state brain functional networks from fMRI data of adult ADHD and healthy controls and found large-scale brain network changes in adult ADHD patients, such as enhanced connectivity between executive control network and right frontoparietal network. In addition, we found that older ADHD spent more time in the pattern of relatively weak connectivity. These findings indicate that the method can effectively extract large-scale functional networks and provide new insights into understanding the neurobiological mechanisms of adult ADHD from the perspective of brain networks.

Keywords: attention deficit hyperactivity disorder, constrained non-negative matrix factorization, dynamic functional connectivity, functional magnetic resonance imaging, intrinsic reference

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a mental disorder portrayed by inattention, hyperactivity, and impulsivity. Studies had shown that approximately 5% of school-age children have ADHD, and symptoms can last to adulthood (Kooij et al., 2010). Considering that children with ADHD are not sensitive to their mental state, it is easier to explore the mystery of ADHD in adults than in children. At present, the etiology and pathogenesis of adult ADHD are not well understood, but it is certain that adult ADHD is inextricably related to the disorder of brain cognitive neural network connectivity. Generally, a healthy brain can be described as an optimized network organization, which is composed of spatially separated brain functional networks with dense functional connectivities (FCs) within brain networks and sparse FCs between brain networks. Brain functional networks play a crucial role in maintaining the balance between functional specialization and integration, supporting individual cognitive, and behavioral abilities (He et al., 2009; Crossley et al., 2013; Bertolero et al., 2015; Sporns and Betzel, 2016). For example, studies had confirmed that the intra-network and inter-network FCs of the frontal parietal network are disrupted in patients with ADHD (Cortese et al., 2012). Accordingly, it is necessary to study ADHD's brain dysfunction network connection in order to identify biomarkers associated with cognitive impairment in patients for clinical diagnosis and optimization of disease treatment, while helping to understand and identify the neurobiological mechanisms of ADHD.

Recent advances had manifested that the combination of resting-state functional magnetic resonance imaging (fMRI) technique and FC analysis framework can help elucidate the disruption of internal brain functional networks in patients with psychiatric disorders (Buckholtz and Meyer-Lindenberg, 2012; Bertolero et al., 2015; Gong et al., 2017). The calculation of FC from the fMRI time series, particularly in the course of the resting state, has manifested a mountain of understanding and awareness concerning the macroscopic spatiotemporal organization of the brain. Moreover, FC within the brain probably varied by psychotic disorders or in the course of evolutionary phases. The secondary analysis of Castellanos et al. (2008) revealed an ADHD-related FC descent pattern between the precuneus and other default mode network components. In addition, previous studies had proved that FC between brain networks changed with increasing age (Allen et al., 2011; Qin et al., 2015). The FC mentioned above is commonly referred to as the static FC approach, assuming that neural activity between brain regions remains stationary throughout the fMRI scan (Calhoun et al., 2014). However, this assumption of stability may not hold because there are no restrictions on brain activity in resting states, and spontaneous fluctuations in brain activity are significant. Taking this property into account, dynamic functional connectivity approaches have been proposed, which can provide unique temporal information (Allen et al., 2014). In addition, brain networks have characteristics that change over time. Hence, this study uses dynamic FC to delve into the interactions within and between large-scale brain networks of ADHD over a period of time, which facilitates the discovery of reproducible patterns of brain functional network impairment and reveals connectivity differences between ADHD patients and healthy control (HC) subjects.

At present, there are many methods to extract brain function networks and their corresponding time series from resting-state fMRI data, such as non-negative matrix decomposition (NMF) and independent component analysis (ICA) (Calhoun et al., 2003; Wang et al., 2004). ICA is a holistic decomposition, where each base is statistically assumed to be independent (Shi et al., 2017, 2018b). However, the presence of multiple zero components in the fMRI data makes it difficult to handle higher-order averages and cannot effectively extract sparse sources. NMF can effectively solve this problem. NMF is a practicable data-driven multi-analysis method that breaks down non-negative data sets into sparse, partially based linear combinations of non-negative features (Deng et al., 2018). In addition, NMF is a nimble approach, which can be developed into a constrained NMF method by adding different constraints to the estimation process. Compared with the classic NMF method, constrained NMF only extracts the expected components by adding prior information to the algorithm, which is beneficial to the subsequent application and reduces the calculation time and storage requirements in the process of NMF analysis. However, the prior information contained in the current constrained NMF method is mostly some specific knowledge about the source signal, such as the experimental paradigm in specific cognitive tasks (Ferdowsi et al., 2010), which does not exist in most cases. Therefore, it is worth considering how to pick up the available authentic information from fMRI data. A recently published article by Shi et al. (2015) put forward a method called group ICA with the intrinsic reference, which extracts the procurable prior message from the group data. Inspired by this method, we presented a new spatial constrained non-negative matrix decomposition (SCNMF) approach to obtain *a priori* information by extracting available real information from the group fMRI data and integrate this *a priori* information into the estimation process of the NMF method.

Generally speaking, NMF is mostly used to analyze fMRI data of a single subject, but in practical application, we need to analyze multiple subjects' fMRI data in more cases. According to the different hypothesis conditions for multi-subjects, there are different processing methods for multi-subject fMRI data analysis, containing spatial concatenation, temporal concatenation, and tensor concatenation. Supposing that total subjects have the same temporal model message, and the fMRI data of multiple subjects are connected according to the spatial dimension, this method is called the spatial concatenation method (Svensén et al., 2002). Supposing that all subjects have the same spatial model message, and the fMRI data of multiple subjects are connected according to the temporal dimension, this method is called the temporal concatenation method (Calhoun et al., 2002; Beckmann et al., 2009). Supposing that all subjects have the same spatial and temporal pattern message except for the different magnitude, and the fMRI data of multiple subjects are connected according to a separate third dimension, this method is called the tensor concatenation method (Beckmann and Smith, 2005; Lee et al., 2008). Previous studies have enunciated that the

temporal concatenation method has a better fMRI group analysis function in these three concatenation methods (Schmithorst and Holland, 2004; Guo and Pagnoni, 2008). Therefore, this manuscript applies the time concatenation way to fMRI data on multiple subjects. At the level of multi-subject, it is simply concatenation according to temporal, which has the problems of a large amount of data sample and high time complexity. To solve this problem, we first reduced the data, then conducted SCNMF analysis on the reduced data, and then reversed reconstruction to obtain the brain functional network information of each subject and its corresponding time process.

In conclusion, in order to better study brain network disorders in adult ADHD patients, we developed a new SCNMF method to extract large-scale functional networks. The proposed SCNMF method was quantitatively compared with two classical methods, including NMF and FastICA. Subsequently, we presented a new analysis framework to evaluate the variations of large-scale functional networks in adult ADHD and HCs. Specifically, the SCNMF method proposed in this manuscript is used to extract the brain function network information of each subject and its corresponding time process, and then dynamic FC and clustering were performed on the time series of all subjects to get the dynamic functional states. The framework calculates group-level connection states by analyzing the dynamic FCs of subjects within a group and then estimates mutually independent subject-specific connection states accordingly, guided by the group-level states. Thus, the resulting subject-specific states capture both inter-subject variability and intra-group similarity. In addition, we used univariate analysis to explore the effect of age on dynamic functional states in adult ADHD and HCs.

MATERIALS AND METHODS

Participants

In this study, the resting fMRI data of 25 adult ADHD and 24 age-matched HCs were obtained from the open database.[1] Demographic characteristics are shown in **Table 1**. The patients and HCs were recruited by the adult ADHD program of the University of New York School of Medicine, whose data collection process and parameters were described as follows: the patient lay flat in the MRI, opened his eyes, remained relaxed, and awoke. The scanning repetition time TR was 2,000 ms, echo time TE was 25 ms, the scanning resolution was 64 × 64, the chip resolution was 3 mm × 3 mm, the slice thickness was 3 mm, the number of slices was 39 slices, covering the whole brain area, and the acquisition consists of 192 time points.

[1]http://www.nitrc.org/projects/fcon_1000/

TABLE 1 | Demographic characteristics.

	ADHD	HC
Age (mean ± SD)	34.75 ± 9.71	34.8 ± 7.99
Sex (male/female)	19/5	15/9

Data Preprocessing

To stabilize the magnetic field of the scanning machine and adapt the subjects to the scanning environment, the first ten time points of the fMRI image were abandoned. The next preprocessing procedures were slice timing and head motion realign. During head motion realign, one ADHD subject was excluded due to excessive head movement, based on a criterion of 3 mm or 3°. Next, the images were registered to the standard template of the Montreal Institute of Neurology (MNI) and then resampled to 2 mm isotropic resolution. Spatial smoothing with a full-width Gaussian core at half-maximum (FWHM) of 6 mm was used to improve the signal-to-noise ratio (SNR) of fMRI data. To reduce noise caused by head movements, the covariate Friston-24 head movement parameter was regressed from the BOLD time series of all voxels.

Non-negative Matrix Factorization

Given the matrix $X \in R_+^{T\times M}$, find the non-negative matrix $W \in R_+^{T\times K}$ and the non-negative matrix $H \in R_+^{K\times M}$, so that:

$$X_{T\times M} \approx W_{T\times K} H_{K\times M} \tag{1}$$

It is approximate because the current solution is not an exact solution, but only a numerical approximation. K is the number of components in the decomposition. The inequality $(T + M)K < T \times M$ represents that only a small number of bases is used to describe a large amount of data. Therefore, it is possible to make $X = WH$ only if W contains the intrinsic characteristics of the random variable.

In addition, NMF is often interpreted by scholars with the background of blind signal separation as a generation model with noise-containing items (Cichocki et al., 2006), which is defined as:

$$X = WH + \varepsilon \tag{2}$$

ε is the noise matrix of $T \times M$. The different NMF algorithms are interpreted as complying with the maximum likelihood algorithm under different ε distribution hypotheses. The process of realizing NMF is a process of optimization solution. Donoho and Stodden (2004) analyzed the requirements for the presence of a unique solution to NMF theoretically. The harshness of this condition revealed that it is a feasible method to construct a reasonable objective function and then optimize W and H alternately to obtain a local optimal solution of NMF. The classic method to seek W and H is by minimizing the discrepancy among X and WH, which is defined as:

$$\min_{W\in R_+^{S\times K}, H\in R_+^{K\times M}} J(W, H) = ||\mathrm{X}-\mathrm{WH}||_F^2 \tag{3}$$

where $||\cdot||_F$ is the Frobenius norm. The matrices W and H are randomly initialized, and then the Euclidean distance function is minimized by updating W and H alternately. The multiplication update rules [30] for W and H are as follows:

$$W_{ij} \leftarrow W_{ij} \frac{(XH^{\mathrm{T}})_{ij}}{(WHH^{\mathrm{T}})_{ij}} \tag{4}$$

$$H_{ij} \leftarrow H_{ij} \frac{(W^{\mathrm{T}}X)_{ij}}{(W^{\mathrm{T}}WH)_{ij}} \tag{5}$$

Extraction of Intrinsic Prior Information

Inspired by the group ICA with intrinsic reference proposed by Shi et al. (2015), we presented an NMF with intrinsic reference method based on the intrinsic reference signal to guide the NMF analysis of fMRI data by mining the available prior information from multiple subject fMRI data. The general process framework for this method is shown in **Figure 1A**.

Assume that there are N subjects in a group fMRI data set, and each subject after normalization has T time points and M voxels. First, NMF analysis is performed separately for each subject, and for subject i, NMF is expressed as:

$$X_i = W_i H_i \ \ (i = 1, 2, \ldots, N) \tag{6}$$

Where X_i represents the $T \times M$ fMRI observation data, W_i represents the $T \times K_i$ mixing matrix, and each row of $H_i = \left(H_i^1, H_i^2, \ldots, H_i^{k_i}\right)^{\mathrm{T}}$ represents a component of subject i after NMF.

We selected ten resting-state functional network templates extracted from the BrainMap database (Smith et al., 2009) as the region of interest (ROI), including the medial visual network (MVISN), occipital pole visual network (OVISN), lateral visual network (LVISN), default mode network (DMN), cerebellum network (CBN), sensorimotor network (SMN), auditory network (AUDN), executive control network (ECN), right frontoparietal network (RFPN), and left frontoparietal network (LFPN). The visualization of the ten resting-state functional networks is shown in **Figure 2**. According to the correlation coefficient between the network template and the components of each subject, the component with the maximum correlation coefficient was selected and denoted as $H_i^{k_i} (i = 1, 2, \ldots, N)$ (6)

Subsequently, we extracted the spatial prior information from $H_i^{k_i} (i = 1, 2, \ldots, N)$ by principal component analysis method to reduce the impact of noise (Langley et al., 2010). All $H_i^{k_i} (i = 1, 2, \ldots, N)$ can form a matrix of $N \times M$, namely,

$$\overline{H} = \left(H_1^{k_1}, H_2^{k_2}, \ldots, H_N^{k_N}\right)^{\mathrm{T}} \tag{7}$$

Whereafter, the eigenvalues and eigenvectors of the covariance matrices of $\overline{H}$ can be calculated. The eigenvector of $\overline{H}$ is expressed as $e_n (n = 1, 2, \ldots, N)$. Furthermore, the first principal component $R = e_1^T \overline{H}$ can be obtained, which is the implicit spatial prior information related to the component of interest in the fMRI data of the subjects in the group, where e_1 represents the feature vector corresponding to the maximum eigenvalue.

Spatial Constrained Non-negative Matrix Factorization

When the prior information is obtained by the above method, we formulate a spatial constrained NMF method containing spatial prior information grounded on the multi-objective optimization framework as follows:

$$\begin{aligned} &Minimize \begin{cases} J(W, H) = ||\mathrm{X–WH}||_F^2 \\ \varepsilon(H) = ||\mathrm{H–R}||_F^2 \end{cases} \\ &Subject\ to\ W \geq 0,\ H \geq 0 \end{aligned} \tag{8}$$

where R represents the spatial prior signal and $\varepsilon(H) = ||\mathrm{H\text{-}R}||_F^2$ is used to measure the similarity between H and R.

Therefore, the objective function $f(W, H)$ can be redefined as follows:

$$f(W, H) = a \cdot J(W, H) + b \cdot \varepsilon(H) \tag{9}$$

where a and b are the weighting parameters.

$$J(W, H) = tr\left(X^{\mathrm{T}} X\right) - 2tr\left(H^{\mathrm{T}} W^{\mathrm{T}} X\right) + + tr\left(H^{\mathrm{T}} W^{\mathrm{T}} W H\right) \tag{10}$$

$$\varepsilon(H) = tr\left(H^{\mathrm{T}} H\right) - 2tr\left(R^{\mathrm{T}} H\right) + tr\left(R^{\mathrm{T}} R\right) \tag{11}$$

Let ψ_{ij} and φ_{ij} be constrained Lagrange multipliers of $W_{ij} \geq 0$ *and* $H_{ij} \geq 0$:

$$L(W, H) = f(W, H) + tr\left(\Psi W^{\mathrm{T}}\right) + tr\left(\Phi H^{\mathrm{T}}\right) \tag{12}$$

where $\Psi = \left[\psi_{ij}\right],\ \Phi = \left[\varphi_{ij}\right]$

The partial derivatives of $L(W, H)$ with respect to W and H are:

$$\frac{\partial L}{\partial W} = -2a \cdot XH^{\mathrm{T}} + 2a \cdot WHH^{\mathrm{T}} + \ \Psi \tag{13}$$

$$\frac{\partial L}{\partial H} = -2a \cdot W^{\mathrm{T}} X + 2a \cdot W^{\mathrm{T}} W H + 2b \cdot H - 2b \cdot R + \Phi \tag{14}$$

Based on the Karush–Kuhn–Tucker (KKT) condition, $\Psi_{ij} W_{ij} = 0$ *and* $\Phi_{ij} H_{ij} = 0$. We can obtain the equation of W_{ij} and H_{ij}:

$$\left(XH^{\mathrm{T}}\right)_{ij} W_{ij} + \left(WHH^{\mathrm{T}}\right)_{ij} W_{ij} = 0 \tag{15}$$

$$a \cdot \left(-W^{\mathrm{T}} X + W^{\mathrm{T}} W H\right)_{ij} H_{ij} + b \cdot (H - R)_{ij} H_{ij} = 0 \tag{16}$$

Whereafter, the update rules of W and H are as follows,

$$W_{ij} \leftarrow W_{ij} \frac{\left(XH^{\mathrm{T}}\right)_{ij}}{\left(WHH^{\mathrm{T}}\right)_{ij}} \tag{17}$$

$$H_{ij} \leftarrow H_{ij} \frac{\left(a \cdot W^{\mathrm{T}} X + b \cdot R\right)_{ij}}{\left(a \cdot W^{\mathrm{T}} W H + b \cdot H\right)_{ij}} \tag{18}$$

Consequently, W and H are constantly updated to meet the convergence rule, that is, to reach the comparative error of the set number or the iteration number of the set value.

When the SCNMF algorithm proposed above extracts fMRI data of multiple subjects, there are problems of large data sample size and time redundancy. Therefore, in this study, the fMRI data of a group of subjects were first downscaled, and then SCNMF analysis was performed on the downscaled data, followed by inverse reconstruction to obtain the brain functional network information of each subject and its corresponding time series, as shown in **Figure 1A**.

Data reduction is to cut down the temporal dimensionality of fMRI data for the benefit of subsequent applications. Two data reduction steps are used for multiple subjects. Supposing that there are N subjects in the aggregate, X_i is the $T \times M$ data matrix from subject i, T is the number of fMRI time points, and M is

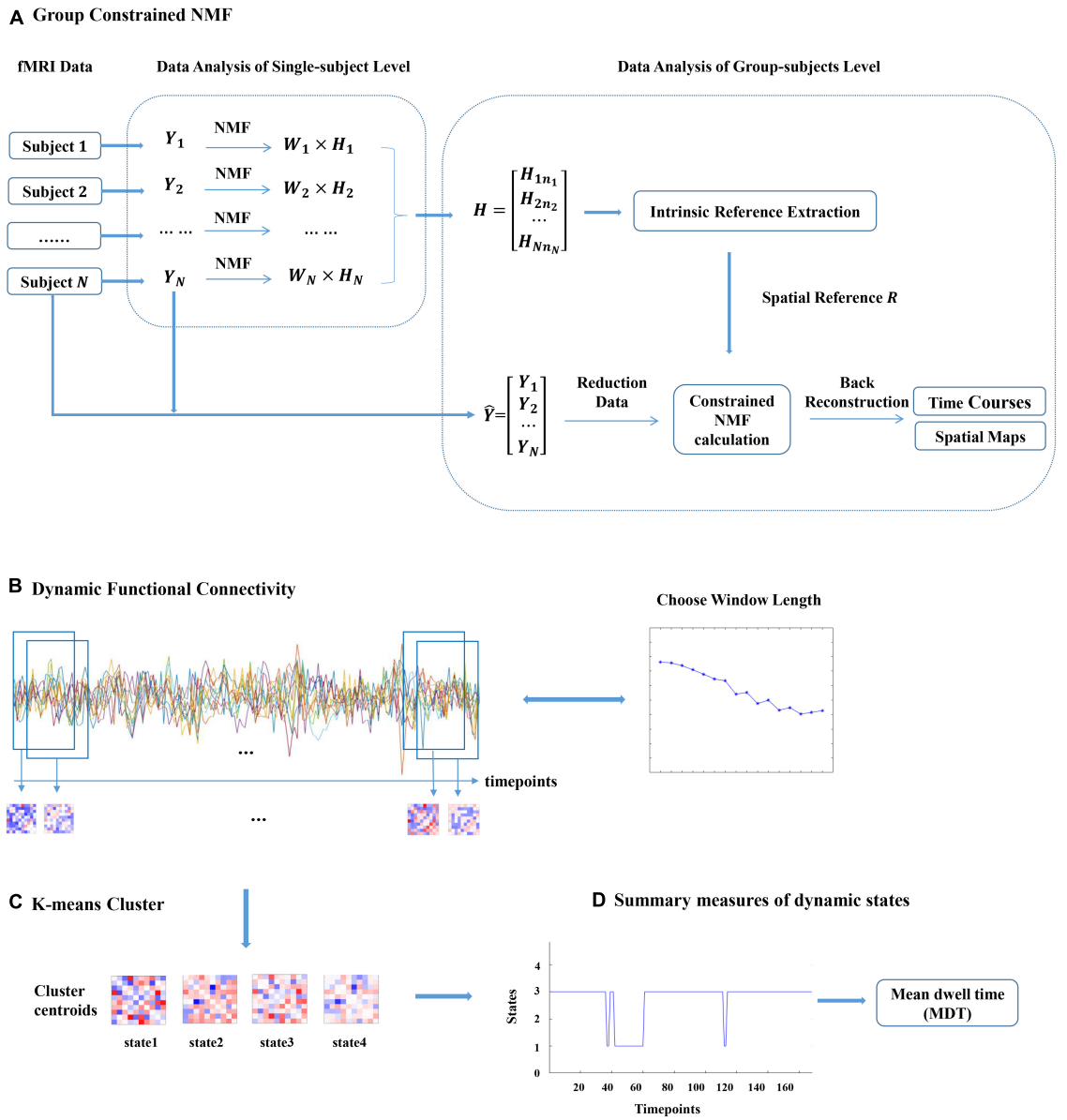


FIGURE 1 | Graphical description of the analysis approach. **(A)** X_i ($i = 1, 2, \ldots, N$) denotes the functional magnetic resonance imaging (fMRI) data of subject i. W_i and H_i denote the temporal and spatial components of subject i gained by non-negative matrix decomposition (NMF). R denotes the spatial prior information obtained from H by the proposed NMF with the intrinsic reference method and incorporates this prior information into the new constrained NMF method. The proposed SCNMF method for the group fMRI analysis is composed of data reduction, constrained NMF calculation, and back reconstruction. **(B)** The sliding window approach is adopted to construct a dynamic FC matrix between each of the selected subject's component. **(C)** K-means clustering is applied to identify dynamic FC states. **(D)** Summary measures of dynamic states, such as the mean dwell time (MDT) and the fraction of time (FT).

the number of voxels. K is the number of estimated components in the temporal dimension. The first step is to perform a PCA descending dimension for each subject, as follows:

$$Y_i = F_i^{-1} X_i \tag{19}$$

where F_i^{-1} is the $L \times T$ reducing matrix. After the fMRI data of each subject is reduced, they are concatenated into a matrix according to the temporal dimension and then used in the next data reduction step. Let Y denote the $S \times M$ reduced data matrix, where S is the temporal dimension after the second dimensionality reduction. Then, the concatenated matrix for N subjects can be obtained as follows:

$$Y = G^{-1} \begin{bmatrix} F_1^{-1} X_1 \\ \cdots \\ F_N^{-1} X_N \end{bmatrix} \tag{20}$$

where G^{-1} is a reducing matrix $S \times LN$.

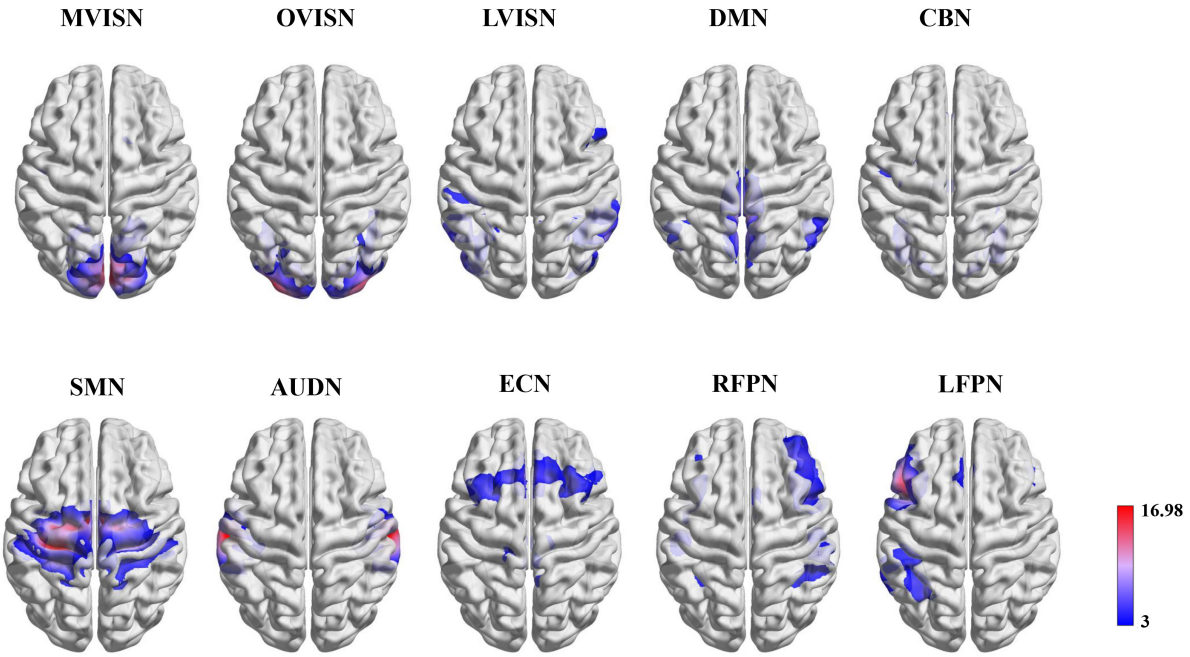


FIGURE 2 | Functional network maps. The functional network z-score maps defined by the ten resting-state functional networks.

Subsequently, constrained NMF is carried out to obtain the $S \times K$ non-negative matrix W and $K \times M$ non-negative matrix H, which is denoted as:

$$Y = WH \tag{21}$$

Then, to calculate the decomposition components of each subject, back reconstruction needs the results of NMF decomposition and data reduction. Subject-specific spatial maps H_i and time courses W_i are gained employing the spatiotemporal regression back reconstruction approach (Erhardt et al., 2011), namely,

$$W_i = X_i H^{-1} \tag{22}$$

$$H_i = W_i^{-1} X_i \tag{23}$$

The matrix H_i contains the K spatial maps and the matrix W_i is the time series matrix of subject i, which consists of the time point corresponding to K components.

In addition, to evaluate the decomposition performance of the SCNMF method, SCNMF was quantitatively compared with two traditional methods, including NMF and FastICA. Specifically, the brain networks of interest were obtained from fMRI data using the SCNMF, NMF, and FastICA methods, respectively. Subsequently, the correlation analysis of the brain network of interest with ten resting-state network templates was carried out, and the experiment was repeated 20 times.

Immediately afterward, the generalized estimating equation method was used to explore whether the correlation coefficients between brain functional templates and brain network components obtained by the three methods (i.e., SCNMF, NMF, and FastICA) were statistically different. Paired *post hoc* analyses were performed for the three methods with Bonferroni correction for multiple comparisons.

Dynamic Functional Connectivity Analysis

To verify that the proposed method SCNMF in this manuscript can effectively extract large-scale functional networks associated with cognitive deficits, we applied the SCNMF method to fMRI data of adult ADHD and HCs, and mainly analyzed the dynamic FC between the large-scale networks of the two groups of subjects. The sliding window approach was adopted to construct a dynamic FC matrix between ten networks (see **Figure 1B**). Starting from the first time point, the time series of brain networks were intercepted with a certain window length in steps of 1 TR. Then Pearson correlation coefficients were calculated for the time series within each window length.

After obtaining the dynamic FC matrices of adult ADHD and HCs, respectively, the combination of these dynamics FC of the two groups was regarded as the cluster samples, and then they were used for k-means clustering (see **Figure 1C**). However, the traditional k-means algorithm has randomness in centroid initialization. Yao et al. (2013) indicated that automatic target generation processing (ATGP) could effectively obtain the initial mixing matrix in the ATGP-ICA algorithm so that ATGP-ICA has better performance than the traditional random initialization FastICA algorithm in the analysis of fMRI signal. Besides, the sparse dictionary learning separation (SDLS) model proposed by Wang et al. (2016) also manifested that ATGP initialization can promote the convergence speed and correctness of SDLS in the process of sparse dictionary learning. Accordingly, the ATGP algorithm was applied to seek effective initial centroid from

dynamic FC matrix to assist the k-means algorithm in achieving stable clustering results. Furthermore, the optimum value of centroid states was taken stock of, applying the elbow criterion defined as the proportion of internal to among cluster distance.

Then, we used the dynamic state summary metrics mean dwell time (MDT) and time fraction (FT) to investigate the time spent by each subject in different states, as shown in **Figure 1D** (Shi et al., 2018a). Furthermore, while multiple aspects of functional connectivity show potential for clinical application, the utility of brain network status assessment as a reliable tool depends on the ability to interpret abnormal outcomes. Underlying factors, such as age, are expected to have a large impact on functional connectivity (Allen et al., 2011; Rashid et al., 2018). Therefore, we performed the univariate analysis of MDT and FT in each state to investigate the relevance between FC and age in different states. The univariate model was defined as follows:

$$\tilde{K} = D\beta \tag{24}$$

where β is the linear regression coefficient and D represents the age matrix for all subjects. Here, age is log-transformed.

RESULTS

Algorithm Decomposition Performance Analysis

In this study, W and H are initialized by assigning values to uniform distributions ranging from 0 to 1. By way of verifying the repeatability of the model, the algorithm model was repeatedly run 20 times in this study to calculate the overlap rate of all results. All the elements of the ten brain network components obtained from the decomposition were to do a union operation to obtain a one-dimensional eigenvector.

The overlap rate is defined as:

$$O = \frac{length\left(union\left(Q_i, Q_j\right)\right)}{\min\left(length(Q_i),\ length\left(Q_j\right)\right)} \tag{25}$$

where Q_i and Q_j, respectively, represent the eigenvectors of the ith and jth results, *length*() represents the number of eigenvectors, and *union*() represents the union of the ith and jth results.

As displayed in **Figure 3A**, the relative error of the objective function only changes slightly after multiple runs of the algorithm. Moreover, in **Figure 3B**, each scatter represents the overlap rate of two results, and it can be found that the overlap rate of features repeated 20 times is more than 0.85. Hence, these pieces of evidence prove the robustness and repeatability of the SCNMF algorithm.

Comparison With Classical Methods

To further verify the effectiveness of the SCNMF method proposed in this manuscript, the correlation analysis between the components obtained by the SCNMF, NMF, and FastICA methods and the ten resting-state network templates obtained by Smith was performed in this study. The experiment was repeated 20 times. **Figure 4** shows the histograms of the correlation coefficients between the components calculated by SCNMF, NMF, and FastICA methods and the ten network templates, respectively. From **Figure 4**, it can be seen that among the ten brain networks, the correlations between the components obtained by SCNMF and the brain network templates were higher than those obtained by NMF and FastICA.

The generalized estimating equation method was used to investigate whether there was a statistical difference in the correlation coefficients between the brain network templates and brain network components obtained by the three methods. The results showed that there was a significant difference between the three methods ($p < 0.001$). Subsequently, paired *post hoc* analyses were performed for the three methods with Bonferroni correction for multiple comparisons, and the results are shown in **Table 2**. As can be seen from the table, there were significant differences between any two of the three methods. This indicates that the SCNMF method proposed by this study does help to improve the spatial source signal recovery in fMRI data analysis and better extract the region of interest from fMRI data compared with the classical NMF and FastICA methods.

Selection of Window Width and Step Length

At present, there is no uniform regulation on the selection of time window length when using the sliding window method. According to previous studies, when using the sliding window for dynamic FC analysis, the cognitive state can be correctly identified when the window size is 30–60 s (Shirer et al., 2012). To more validly capture the dynamic pattern of psychiatric disorders and HCs, we conducted softmax classification for these four groups of subjects and determined the optimal window width through the highest classification rate. First, we calculated the correlation matrix of each subject between 15TR and 30TR, and the step size was 1TR. Each correlation matrix was triangulated to form a one-dimensional vector (45×1) as a classification feature. Repeat the above two steps to calculate the classification characteristics of all the subjects. Whereafter, the characteristics of these four subjects were classified by softmax. The ratio of training samples and test samples was 7:3. As displayed in **Figure 5A**, when the window width was 22TR, the classification rate was the highest. Therefore, we chose the optimal window width of 29TR for the dynamic FC of all subjects.

Meanwhile, we used the same softmax classification algorithm to determine the step size of the sliding time window and calculated the classification rate for step sizes from 1TR to 10TR. From **Figure 5B**, it can be seen that the classification rate was high when the step size is 5TR, 7TR, 8TR, 9TR, and 10TR, while referring to the results of existing studies, 5TR was chosen as the step size in this study.

Dynamic Connectivity States

In this study, the k-means clustering method was used to cluster the dynamic connectivity matrix of all subjects defined as input samples. In addition, the elbow criterion was used to determine the optimal clustering performance, i.e., the number of clusters was set to 4. By k-means clustering on account of the dynamic FC window, four states that appear repeatedly in

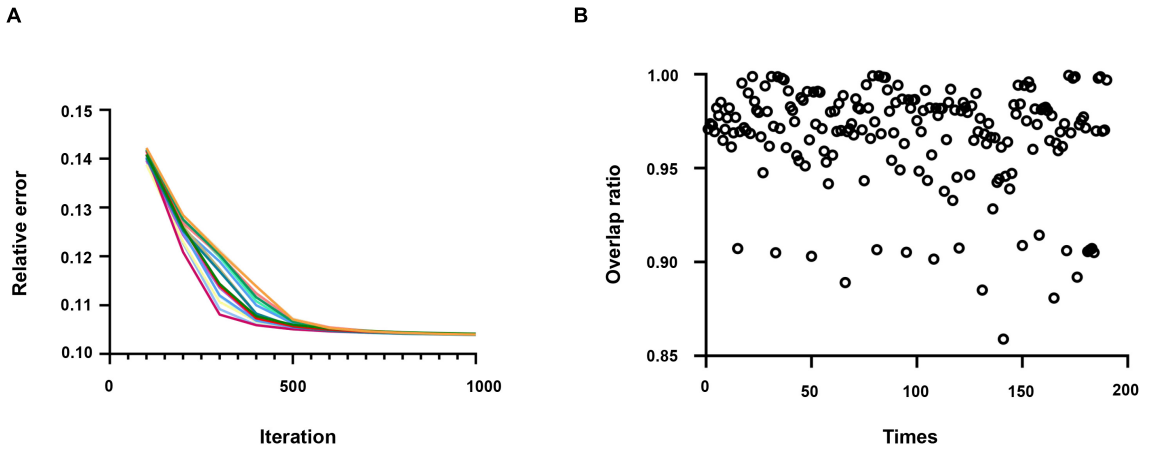


FIGURE 3 | (A) The relative error of the results repeated running 20 times. Each time, the result is indicated by a different color. **(B)** The overlap rate of the results was repeated 20 times.

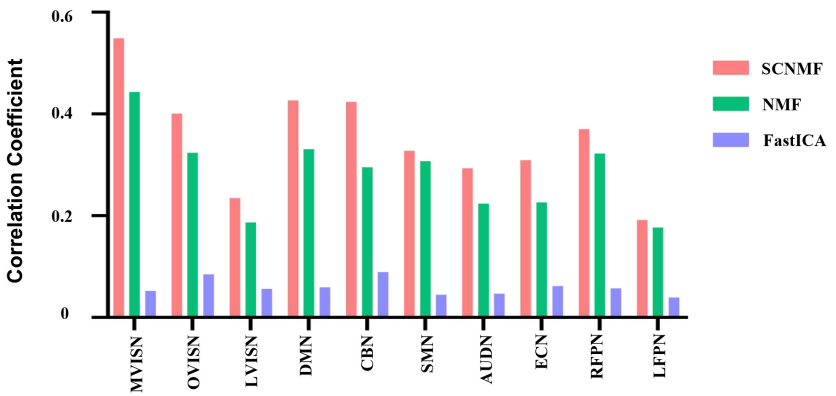


FIGURE 4 | The proposed algorithm is compared with NMF and FastICA. SCNMF, NMF, and FastICA are used to extract large-scale brain network components, and the correlation coefficients between components and network templates were calculated, respectively.

the whole scanning process and across subjects were obtained. The specific distribution of adult ADHD and HCs dynamic windows in different states is shown in **Figure 6**. As can be seen from **Figure 6**, adult ADHD was mainly distributed in state 2 and state 4, and HC subjects were concentrated in state 1 and state 3. Interestingly, each state had a representative subject that accounted for a large proportion of subjects, with 84.47% of HC subjects in state 1, 26.13% of ADHD subjects in state 2, 13.38% of HC subjects in state 3, and 73.36% of ADHD subjects in state 4.

What is more, **Figure 7A** shows the four dynamic FC states for adult ADHD and HCs. Each state represented the centroid of a state and presumptively reflected a connectivity state steadily presented in the fMRI data. A distinctive feature was that the overall connectivity of state 1 to state 4 was different in terms of the strength of the connectivities between the functional networks. Specifically, state 2 and state 3 were stronger connected states, and state 1 and state 4 were weaker connected states. It was worth noting that the weaker connectivity state was easier to detect than the stronger connectivity state, which meant that the connectivity of the brain network mainly depended on the comparatively weaker connectivity state.

To better visualize the connectivity patterns of the large-scale network in different states, **Figure 7B** displays the top 15% of the absolute value of the connectivity strength in each state. In states 1 and 3, the large-scale network exhibited connectivity characteristics around LVISN and OVISN, and also involved RFPN. Furthermore, the connectivity between SMN and RFRN was positively connected in both states. On the whole, the connectivity distribution was more complex. In state 2, the large-scale network presented connectivity characteristics around DMN and RFPN. In state 4, the overall connectivity strength

TABLE 2 | Multiple comparisons.

				95% confidence interval	
Method	**Method**	***df***	**Sig**	**Lower bound**	**Upper Bound**
SCNMF	NMF	1	$p = 0.034$	0.0052	0.1871
	FastICA	1	$p < 0.001$	0.2242	0.3795
NMF	FastICA	1	$p < 0.001$	0.1549	0.2563

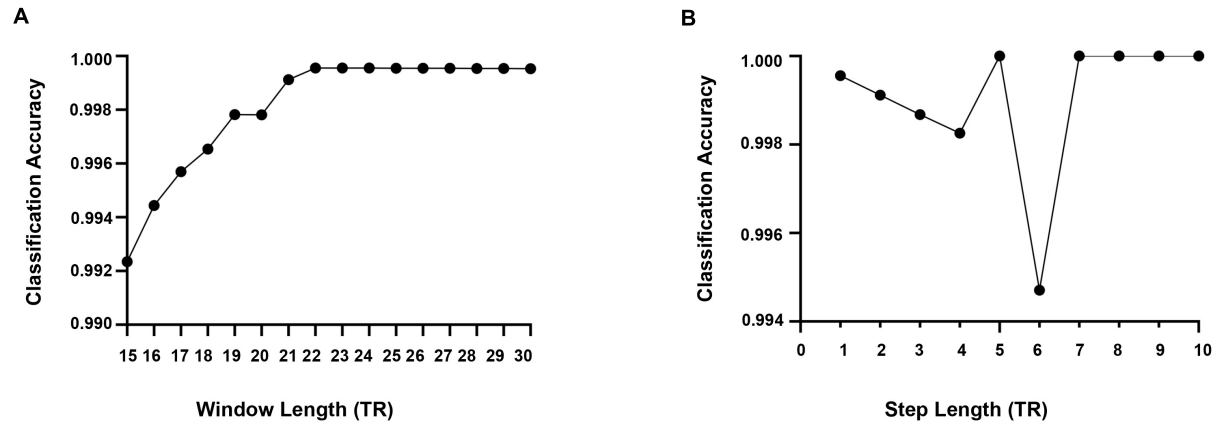


FIGURE 5 | (A) The optimal window length is estimated based on the classification accuracy. **(B)** Estimation results of the optimal step length in terms of classification accuracy.

was weaker. At the same time, the overall connectivity strength of adult ADHD patients was lower than that of HCs, indicating that the connectivities between large-scale networks of adult ADHD subjects were interrupted or decreased. In addition, this illustrates that the SCNMF algorithm can be effective in detecting patterns of connectivity specific to the large-scale networks of adult ADHD patients and HCs.

To explore how different quantitative summary metrics of dynamic FC vary with age, the MDT for each state is statistically analyzed employing univariate linear regression, and the results are shown in **Figure 8**. The parameter β stands for the linear regression coefficient. A positive coefficient of β expresses that the older subjects expended more time in the commensurable state whereas a subtractive coefficient of β indicated that younger subjects expended more time in that state.

As can be seen in **Figure 8**, older subjects had longer MDT in state 1, state 3, and state 4. In contrast, younger subjects showed longer MDT in state 2. Since adult ADHD was mainly distributed in states 2 and 4, it can be seen that older adults with ADHD were mainly in state 4 and younger adults with ADHD were mainly in state 2. However, the effect of age was smaller for HC subjects.

To further research the influence of age on adult ADHD in states 2 and 4, the median participant in each state was calculated employing the window correlation matrix of each subject. It has been documented that the median matrix can be deemed a representative model of the subjects in a specific state (Damaraju et al., 2014). Subsequently, univariate analysis was applied to research the correlation between connectivity and age in the median matrix of these subjects, as shown in **Figure 9A**.

The effect of age on the large-scale network connectivity of adult ADHD subjects in state 2 and state 4 was only reflected in a few large-scale network connectivities. However, the effects of age on the two states were different. Specifically, in state 2, the negative age-related FC was between CBN and ECN. In state 4, age only affects the FC between OVISN and LFPN.

Then, to assess adult ADHD state differences in dynamic functional connectivity between state 2 and state 4, we used the median of these subjects and applied paired t-test to assess state differences. All outcomes were consistent with the false discovery rate (FDR) multiple comparison correction threshold $p < 0.05$. The state modification was visualized by drawing the log of p value with the sign of t statistics, $-sign(t)\log(p)$. The momentous state difference between functional networks is shown in **Figure 9B**. The results showed that FC in adult ADHD was significantly different under these two states. Among them, the positively significant differences in brain functional connectivity were mainly between ECN and CBN, and between AUDN and SMN. Relatively speaking, the FCs with negatively significant differences were mainly between LVISN and SMN. These pieces of evidence indicate that large-scale network connectivities in adult ADHD have commonality and specificity and may serve as a biomarker for the study of adult ADHD in the future.

DISCUSSION

In this study, we proposed a new spatial constrained NMF method based on the authentic reference signal, which mined available spatial prior information from the group fMRI

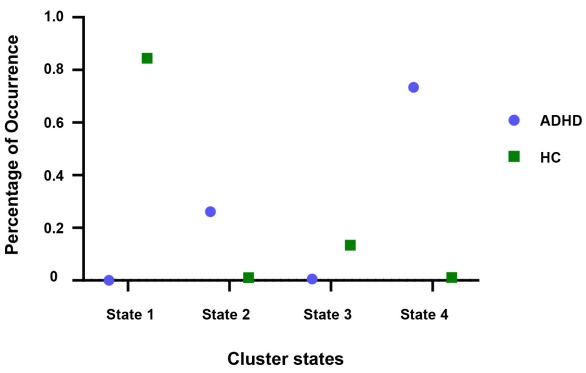


FIGURE 6 | Percentage of recurring dynamic windows for adult ADHD and HCs in each state.

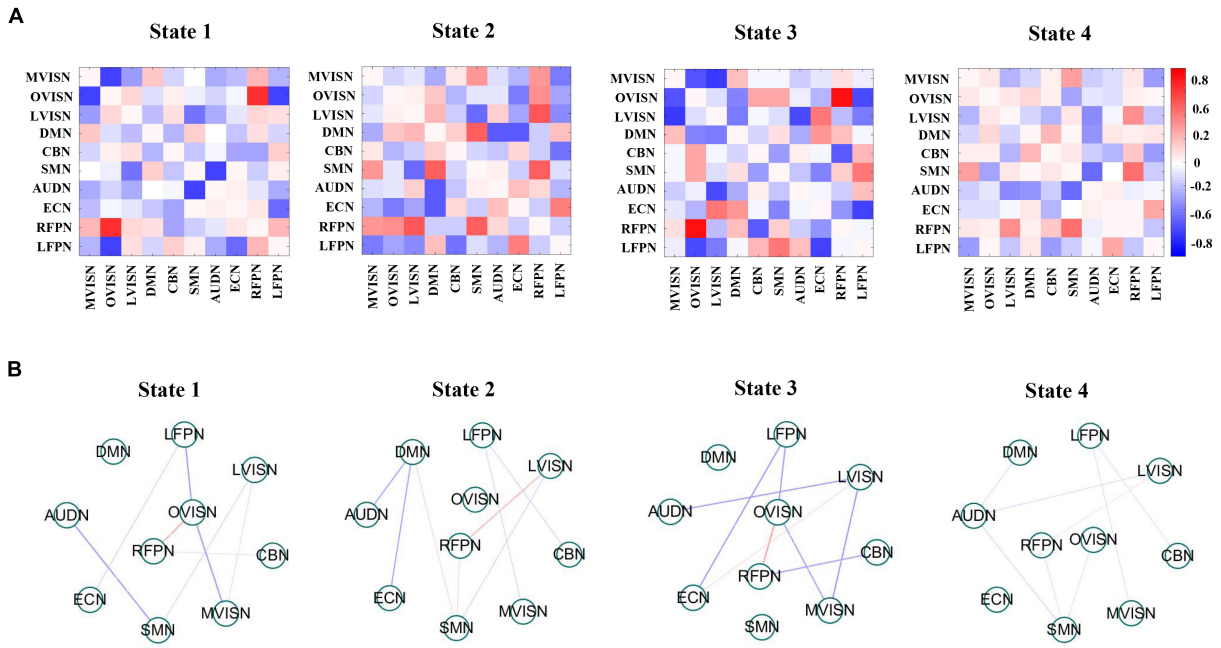


FIGURE 7 | (A) Group-specific centroids of dynamic functional connectivity (FC) states for adult Attention-deficit/hyperactivity disorder (ADHD) and healthy controls (HCs). Dynamic FC states stemmed from k-means clustering employing the sliding window approach. **(B)** It shows that the absolute strength of dynamic FC of large-scale networks in each state is at the top 15%. The red line expresses positive connectivity between networks, and the blue line expresses minus connectivity between networks. The thickness of the line indicates the strength of connectivity.

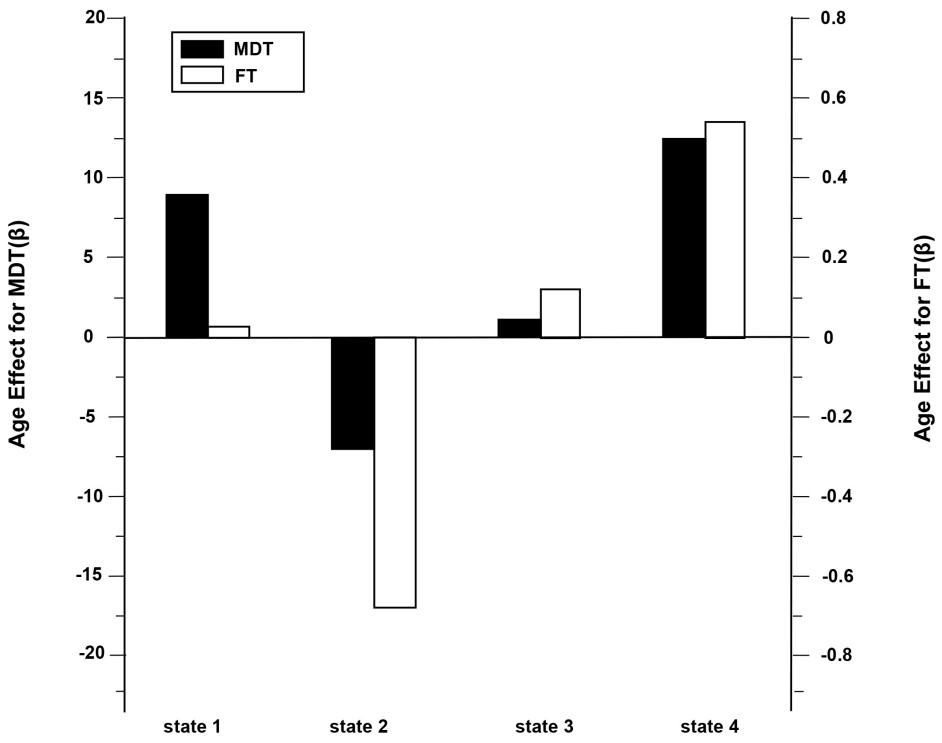


FIGURE 8 | The mean dwell time (MDT) and fraction of time (FT) from four states varied with age. The positive coefficient of β manifests that the older subjects expend more time in the homologous state. Simultaneously, the subtractive coefficient of β manifests that the younger subjects expend more time in that state.

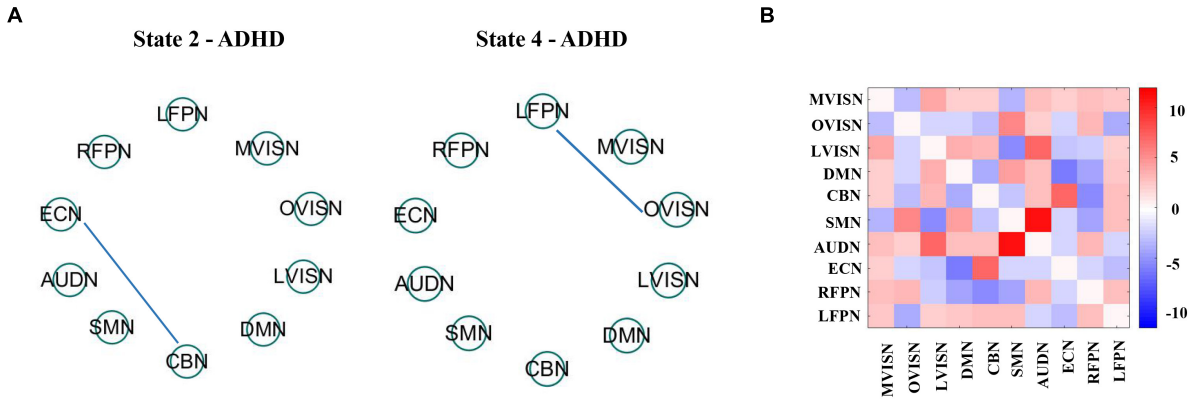


FIGURE 9 | (A) It shows age association across the dynamic FC state 2 and state 4. The blue line expresses a negative relevance among FC and age. All of the outcomes expressed correspond to the false discovery rate (FDR) multiple comparison correction threshold of $p < 0.05$. **(B)** State difference between dynamic FC state 2 and state 4 is obtained using paired *t*-test with the FDR threshold ($p < 0.05$). State variation was visualized by drawing the log of *p* value with the sign of *t* statistics, $-sign(t)\log(p)$.

data and amalgamated this prior information into the NMF method. By comparing with other classical methods, the SCNMF algorithm proposed in this manuscript can better extract the region of interest. On this basis, this study proposed a new analytical framework for assessing large-scale functional network changes in adult ADHD and healthy subjects, exploring the differences in dynamic functional connectivity patterns between the two groups. To our knowledge, this is the first time that NMF has been integrated into the dynamic functional connectivity of large-scale networks and applied to adult ADHD.

So far, the most common dynamic functional connectivity method for fMRI data is ICA. The ICA method assumes independence between components to obtain spatial components, but there is a problem that it cannot extract sparse sources efficiently. The NMF method can be a good solution to this problem because it can decompose the data into a linear combination of sparse, partial-based non-negative features. In addition, it has been shown that adding prior information to NMF can improve its analytical performance on fMRI data. Therefore, this study proposed an SCNMF method based on the real information of fMRI data. The proposed SCNMF method was quantitatively compared with the traditional methods NMF and FastICA, and the results showed that the spatial brain networks detected by SCNMF were better than those detected by NMF and FastICA. From this point of view, it seems that the SCNMF method does help to improve the recovery of spatial source signals in fMRI data analysis, while the extracted brain networks of interest have higher separation quality and accuracy.

Dynamic FC collared steady connectivity patterns, which are not watched in stationary FC. The FC of the brain is not static but changed over time. Therefore, more valuable information can be discovered by observing the differences in intergroup connectivity captured by dynamic states. According to the results of this study, the differences between ADHD and HCs were not limited to a single dynamic FC state but distributed in four dynamic FC states.

The ECN consists of the cingulate cortex, prefrontal cortex, insular cortex, and striatum and is involved in cognition, inhibitory behavior, emotion, and pain (Smith and Nichols, 2009). It is also referred to as the "transitional network connecting cognition and emotion/sensation" (Laird et al., 2011). The medial prefrontal cortex is an important component of the ECN and is an area known to be associated with information processing. In most psychiatric disorders, coordination between executive function and internal and external attention is considered to be severely impaired. Interestingly, ADHD patients had elevated connectivity between the ECN and RFPN in the dynamic functional connectivity state compared to HCs. Furthermore, it had been shown that ADHD abnormalities were strongly associated with the ECN (Makris et al., 2009; Bush, 2010; Posner et al., 2014). Therefore, the unbalanced connectivity between ECN and FPN may help to explain the broader psychopathological features of ECN abnormalities in adult ADHD.

Another important finding of dynamic FC between large-scale networks was that adult ADHD had enhanced FC between ECN and CBN compared to HCs, but connectivity became smaller with age. According to a literature review, ADHD patients had demonstrated stronger internal connectivity in the CBN (Mostert et al., 2016). In addition, an increased FC in the ECN and the anterior cingulate cortex of the cerebellum was previously found in ADHD (McCarthy et al., 2013; Wang et al., 2013), which was consistent with our findings. However, our results complemented existing studies and found out the effects of age on ADHD.

Cortese et al. conducted a meta-analysis of 55 articles (39 for children and 16 for adults), concentrating on clinical features or particular neuropsychological structures in ADHD. They discovered that adult ADHD hypoactivation was prominent in the frontoparietal networks, while adult ADHD hyperactivation appeared in the visual, dorsal attention, and DMN (Cortese et al., 2012). In the present study, adult ADHD had enhanced connectivity in the DMN and SMN relative to HCs. Interestingly, adult ADHD also had enhanced RFPN and SMN connectivity.

Therefore, there was a great difference between the strength of the FC within the network and the FC between the networks. We can take this as a starting point and verify it in the future research center.

It is worth mentioning that there is an age-related effect of MDT in dynamic states in adult ADHD and HCs. Specifically, it was observed that older ADHD had a longer dwell time in the relatively weaker connected state 4, whereas younger ADHD had a shorter dwell time in the relatively stronger connected state 2. However, age had little effect on HCs. This is slightly different from the adult connectivity model previously found in adult studies (Fair et al., 2008; Zuo et al., 2010), suggesting that neurodevelopment in ADHD differs from that of HCs. In addition, the difference in results may also be due to differences in the scale of the study. In this manuscript, the entry point of the study is the brain network, while some literature has considered brain regions, and the difference in scale may also affect the final results. In the future, we can consider more comprehensive studies from multiple scales to further discover the pathogenesis of ADHD.

In addition, there was almost no overlap in the four states obtained in dynamic functional connectivity with only one type of subjects in each state. Considering that it is possible that this is due to a problem with the data set, for this reason, we changed the data set for validation. Forty-five HCs and 40 adult ADHD subjects were downloaded from openfMRI. Ten cognitive networks were extracted using the algorithm proposed in this manuscript, followed by dynamic functional connectivity analysis. The four dynamic functional state distributions obtained were similar to the results obtained in this manuscript.

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

YL and WZ conceived and designed the study. YL performed the experiments and wrote the manuscript. YL, JY, and WN collected the samples and analyzed the data. JD, YS, and SL reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31870979 and 61906117), Shanghai Sailing Program (Grant No. 19YF1419000), and the Science and Technology Support Projects of the Shanghai Science and Technology Committee (Grant No. 19411971400).

REFERENCES

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking Whole-Brain Connectivity Dynamics in the Resting State. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352

Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., et al. (2011). A baseline for the multivariate comparison of resting-state networks. *Front. Syst. Neurosci.* 5:2. doi: 10.3389/fnsys.2011.00002

Beckmann, C., Mackay, C., Filippini, N., and Smith, S. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *Neuroimage* 47, S39–S41.

Beckmann, C. F., and Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *Neuroimage* 25, 294–311. doi: 10.1016/j.neuroimage.2004.10.043

Bertolero, M. A., Yeo, B. T. T., and D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proc.Natl. Acad. Sci.U.S.A.* 112, E6798–E6807. doi: 10.1073/pnas.1510619112

Buckholtz, J. W., and Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. *Neuron* 74, 990–1004. doi: 10.1016/j.neuron.2012.06.002

Bush, G. (2010). Attention-Deficit/Hyperactivity Disorder and Attention Networks. *Neuropsychopharmacology* 35, 278–300. doi: 10.1038/npp.2009.120

Calhoun, V. D., Adali, T., Hansen, L. K., Larsen, J., and Pekar, J. J. (2003). *ICA of Functional MRI Data: An Overview.4th International Symposium*. Switzerland: International cartographic association location.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2002). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 16, 140–51. doi: 10.1002/hbm.10044

Calhoun, V. D., Miller, R., Pearlson, G., and Adali, T. (2014). The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. *Neuron* 84, 262–274. doi: 10.1016/j.neuron.2014.10.015

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., et al. (2008). Cingulate-precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biol. Psychiat.* 63, 332–337. doi: 10.1016/j.biopsych.2007.06.025

Cichocki, A., Zdunek, R., and Amari, S. (2006). Csiszar's divergences for non-negative matrix factorization: Family of new algorithms. *Indep. Component Anal. Blind Signal Sep., Proc.* 3889, 32–39.

Cortese, S., Kelly, C., Chabernaud, C., Proal, E., Di Martino, A., Milham, M. P., et al. (2012). Toward Systems Neuroscience of ADHD: A Meta-Analysis of 55 fMRI Studies. *Am. J. Psychiat.* 169, 1038–1055. doi: 10.1176/appi.ajp.2012.11101521

Crossley, N. A., Mechelli, A., Vertes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., et al. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc.Natil. Acad. Sci.U.S.A.* 110, 11583-11588. doi: 10.1073/pnas.1314559110

Damaraju, E., Allen, E. A., Belger, A., Ford, J. M., McEwen, S., Mathalon, D. H., et al. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *Neuroimage Clin.* 5, 298–308. doi: 10.1016/j.nicl.2014.07.003

Deng, J., Kong, W., Wang, S. Q., Mou, X. Y., and Zeng, W. M. (2018). Prior Knowledge Driven Joint NMF Algorithm for ceRNA Co-Module Identification. *Int.J. Biol. Sci.* 14, 1822–1833. doi: 10.7150/ijbs.27555

Donoho, D., and Stodden, V. (2004). *When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts?*. Amsterdam: *ELSEVIER*.

Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adali, T., and Calhoun, V. D. (2011). Comparison of Multi-Subject ICA Methods for Analysis of fMRI Data. *Hum. Brain Mapp.* 32, 2075–2095. doi: 10.1002/hbm.21170

Fair, D. A., Cohen, A. L., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Barch, D. M., et al. (2008). The maturing architecture of the brain's default network. *Proc.Natl.Acad.Sci.U.S.A.* 105, 4028–4032. doi: 10.1073/pnas.0800376105

Ferdowsi, S., Abolghasemi, V., and Sanei, S. (2010). "A constrained NMF algorithm for bold detection in FMRI," in *Machine Learning for Signal Processing (MLSP)*, (Manhattan, NY: IEEE International). doi: 10.1109/IEMBS.2011.6091251

Gong, Q. Y., Hu, X. Y., Pettersson-Yeo, W., Xu, X., Lui, S., Crossley, N., et al. (2017). Network-Level Dysconnectivity in Drug-Naive First-Episode Psychosis: Dissociating Transdiagnostic and Diagnosis-Specific Alterations. *Neuropsychopharmacology* 42, 933–940. doi: 10.1038/npp.2016.247

Guo, Y., and Pagnoni, G. (2008). A unified framework for group independent component analysis for multi-subject fMRI data. *Neuroimage* 42, 1078–1093. doi: 10.1016/j.neuroimage.2008.05.008

He, Y., Wang, J. H., Wang, L., Chen, Z. J., Yan, C. G., Yang, H., et al. (2009). Uncovering Intrinsic Modular Organization of Spontaneous Brain Activity in Humans. *Plos One* 4:e5226. doi: 10.1371/journal.pone.0005226

Kooij, S. J. J., Bejerot, S., Blackwell, A., Caci, H., Casas-Brugue, M., Carpentier, P. J., et al. (2010). European consensus statement on diagnosis and treatment of adult ADHD: The European Network Adult ADHD. *Bmc Psychiat.* 10:67. doi: 10.1186/1471-244X-10-67

Laird, A. R., Fox, P. M., Eickhoff, S. B., Turner, J. A., Ray, K. L., McKay, D. R., et al. (2011). Behavioral Interpretations of Intrinsic Connectivity Networks. *J. Cogn. Neurosci.* 23, 4022–4037. doi: 10.1162/jocn_a_00077

Langley, P., Bowers, E. J., and Murray, A. (2010). Principal Component Analysis as a Tool for Analyzing Beat-to-Beat Changes in ECG Features: Application to ECG-Derived Respiration. *IEEE Trans. Biomed. Eng.* 57, 821–829. doi: 10.1109/Tbme.2009.2018297

Lee, J. H., Lee, T. W., Jolesz, F. A., and Yoo, S. S. (2008). Independent vector analysis (IVA): Multivariate approach for fMRI group study. *Neuroimage* 40, 86–109. doi: 10.1016/j.neuroimage.2007.11.019

Makris, N., Biederman, J., Monuteaux, M. C., and Seidman, L. J. (2009). Towards Conceptualizing a Neural Systems-Based Anatomy of Attention-Deficit/Hyperactivity Disorder. *Dev. Neurosci.* 31, 36–49. doi: 10.1159/000207492

McCarthy, H., Skokauskas, N., Mulligan, A., Donohoe, G., Mullins, D., Kelly, J., et al. (2013). Attention Network Hypoconnectivity With Default and Affective Network Hyperconnectivity in Adults Diagnosed With Attention-Deficit/Hyperactivity Disorder in Childhood. *Jama Psychiat.* 70, 1329–1337. doi: 10.1001/jamapsychiatry.2013.2174

Mostert, J. C., Shumskaya, E., Mennes, M., Onnink, A. M. H., Hoogman, M., Kan, C. C., et al. (2016). Characterising resting-state functional connectivity in a large sample of adults with ADHD. *Prog. NeuroPsychopharmacol. Biol. Psychiat.* 67, 82–91. doi: 10.1016/j.pnpbp.2016.01.011

Posner, J., Park, C., and Wang, Z. S. (2014). Connecting the Dots: A Review of Resting Connectivity MRI Studies in Attention-Deficit/Hyperactivity Disorder. *Neuropsychol. Rev.* 24, 3–15. doi: 10.1007/s11065-014-9251-z

Qin, J., Chen, S. G., Hu, D. W., Zeng, L. L., Fan, Y. M., Chen, X. P., et al. (2015). Predicting individual brain maturity using dynamic functional connectivity. *Front. Hum. Neurosci.* 9:418. doi: 10.3389/fnhum.2015.00418

Rashid, B., Blanken, L. M. E., Muetzel, R. L., Miller, R., Damaraju, E., Arbabshirani, M. R., et al. (2018). Connectivity dynamics in typical development and its relationship to autistic traits and autism spectrum disorder. *Hum. Brain Mapp.* 39, 3127–3142. doi: 10.1002/hbm.24064

Schmithorst, V. J., and Holland, S. K. (2004). Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *J. Magn. Reson. Imaging* 19, 365–368. doi: 10.1002/jmri.20009

Shi, L., Sun, J. Z., Wu, X. R., Wei, D. T., Chen, Q. L., Yang, W. J., et al. (2018a). Brain networks of happiness: dynamic functional connectivity among the default, cognitive and salience networks relates to subjective well-being. *Soc. Cogn. Affect. Neurosci.* 13, 851–862. doi: 10.1093/scan/nsy059

Shi, Y. H., Zeng, W. M., Wang, N. Z., and Zhao, L. (2018b). A New Constrained Spatiotemporal ICA Method Based on Multi-Objective Optimization for fMRI Data Analysis. *IEEE Trans.act. Neural Syst. Rehabil. Eng.* 26, 1690–1699. doi: 10.1109/Tnsre.2018.2857501

Shi, Y. H., Zeng, W. M., Wang, N. Z., and Chen, D. T. L. (2015). A novel fMRI group data analysis method based on data-driven reference extracting from group subjects. *Comp. Methods Prog. Biomed.* 122, 362–371. doi: 10.1016/j.cmpb.2015.09.002

Shi, Y. H., Zeng, W. M., Wang, N. Z., and Zhao, L. (2017). A new method for independent component analysis with priori information based on multi-objective optimization. *J. Neurosci. Methods* 283, 72–82. doi: 10.1016/j.jneumeth.2017.03.018

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. D. (2012). Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cereb. Cortex* 22, 158–165. doi: 10.1093/cercor/bhr099

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc.Natl. Acad. Sci.U.S.A.* 106, 13040–13045. doi: 10.1073/pnas.0905267106

Smith, S. M., and Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98. doi: 10.1016/j.neuroimage.2008.03.061

Sporns, O., and Betzel, R. F. (2016). Modular Brain Networks. *Annu. Rev. Psychol.* 67, 613–640. doi: 10.1146/annurev-psych-122414-033634

Svensén, M., Kruggel, F., and Benali, H. (2002). ICA of fMRI group study data. *Neuroimage* 16, 551–563. doi: 10.1006/nimg.2002.1122

Wang, N. Z., Zeng, W. M., and Chen, D. T. L. (2016). A Novel Sparse Dictionary Learning Separation (SDLS) Model With Adaptive Dictionary Mutual Incoherence Constraint for fMRI Data Analysis. *IEEE Trans. Biomed. Eng.* 63, 2376–2389. doi: 10.1109/Tbme.2016.2533722

Wang, X., Jie, T., Li, X., Dai, J., and Lin, A. (2004). Detecting brain activations by constrained non-negative matrix factorization from task-related BOLD fMRI. *Proc. SPIE Int. Soc. Opt. Eng.* 5369, 675–682.

Wang, X. H., Jiao, Y., Tang, T. Y., Wang, H., and Lu, Z. H. (2013). Altered regional homogeneity patterns in adults with attention-deficit hyperactivity disorder. *Eur. J. Radiol.* 82, 1552–1557. doi: 10.1016/j.ejrad.2013.04.009

Yao, S. N., Zeng, W. M., Wang, N. Z., and Chen, L. (2013). Validating the performance of one-time decomposition for fMRI analysis using ICA with automatic target generation process. *Magn. Reson. Imaging* 31, 970–975. doi: 10.1016/j.mri.2013.03.014

Zuo, X. N., Kelly, C., Di Martino, A., Mennes, M., Margulies, D. S., Bangaru, S., et al. (2010). Growing Together and Growing Apart: Regional and Sex Differences in the Lifespan Developmental Trajectories of Functional Homotopy. *J. Neurosci.* 30, 15034–15043. doi: 10.1523/Jneurosci.2612-10.2010

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, Zeng, Shi, Deng, Nie, Luo 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 in Neuroscience

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fnins.2022.807585

Abnormal Cerebrovascular Reactivity and Functional Connectivity Caused by White Matter Hyperintensity Contribute to Cognitive Decline

Dan Yang[1], Ruomeng Qin[1], Lan Chu[2], Hengheng Xu[1], Ling Ni[1], Junyi Ma[1], Pengfei Shao[1], Lili Huang[1], Bing Zhang[3], Meijuan Zhang[1,4,5,6] and Yun Xu[1,4,5,6]**

[1] Department of Neurology, Drum Tower Hospital, Medical School and The State Key Laboratory of Pharmaceutical Biotechnology, Institute of Brain Science, Nanjing University, Nanjing, China, [2] Department of Neurology, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [3] Department of Radiology, Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing, China, [4] Jiangsu Key Laboratory of Molecular Medicine, Medical School of Nanjing University, Nanjing, China, [5] Jiangsu Province Stroke Center for Diagnosis and Therapy, Nanjing, China, [6] Nanjing Neurology Clinic Medical Center, Nanjing, China

OPEN ACCESS

Edited by:
Kaijian Xia, Changshu No.1 People's Hospital, China

Reviewed by:
Youngkyoo Jung, University of California, Davis, United States
Yan Zhou, Shanghai Jiao Tong University, China

***Correspondence:**
Meijuan Zhang
juanzi1986126@163.com
Yun Xu
xuyun20042001@aliyun.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *02 November 2021*
Accepted: *10 January 2022*
Published: *04 March 2022*

Citation:
Yang D, Qin R, Chu L, Xu H, Ni L, Ma J, Shao P, Huang L, Zhang B, Zhang M and Xu Y (2022) Abnormal Cerebrovascular Reactivity and Functional Connectivity Caused by White Matter Hyperintensity Contribute to Cognitive Decline. Front. Neurosci. 16:807585. doi: 10.3389/fnins.2022.807585

Aims: This study aimed to investigate the relationships of impaired cerebrovascular reactivity (CVR) and abnormal functional connectivity (FC) with white matter hyperintensity (WMH)-related cognitive decline.

Methods: A total of 233 WMH subjects were recruited and categorized into WMH-I (n = 106), WMH-II (n = 72), and WMH-III (n = 55) groups according to Fazekas visual rating scale. All participants underwent neuropsychological tests and multimodal MRI scans, including 3D-T1, and resting-state functional magnetic resonance imaging (rs-fMRI). The alterations of CVR maps and FC were further explored.

Results: Subjects with a higher WMH burden displayed a lower CVR in the left medial occipital gyrus (MOG). The FC analysis using MOG as a seed revealed that the FC of the left insula, left inferior parietal lobule, and thalamus changed abnormally as WMH aggravated. After adjusting for age, gender, and education years, the serial mediation analysis revealed that periventricular white matter hyperintensity contributes indirectly to poorer Mini-Mental State Examination (MMSE) scores (indirect effect: $\beta = -0.1248$, 95% CI: −0.4689, −0188), poorer Montreal Cognitive Assessment (MoCA) (indirect effect: $\beta = -0.1436$, 95% CI: −0.4584, −0.0292) scores, and longer trail making tests A (TMT-A) (indirect effect: $\beta = 0.1837$, 95% CI: 0.0069, 0.8273) times, specifically due to the lower CVR of the left MOG and the higher FC of the left insula-MOG.

Conclusion: The CVR decline of the left MOG and the abnormal FC of the left insula-MOG attributed to WMH progression were responsible for the poor general cognition (MMSE and MoCA) and information processing speed (TMT-A). The left MOG may act as a connection, which is involved in the processing of cognitive biases by connecting with the left insula-cortical regions in WMH individuals.

Keywords: white matter hyperintensity, cognitive decline, cerebrovascular reactivity, functional connectivity, rs-fMRI

INTRODUCTION

White matter hyperintensity (WMH), presented as hyperintensity of the subcortical white matter on T2-weighted image or fluid-attenuated inversion recovery (FLAIR) sequence, is frequently observed in the brain of elderly individuals (Alber et al., 2019). Epidemiologic studies revealed that WMH can be observed in 72–96% of European individuals over 60 years old (Longstreth et al., 1996) and 70% Chinese individuals over 50 years old (Han et al., 2018). Increasing evidences have revealed that the WMH burden exerts a negative impact on cognition, mainly in executive function and information process speed (Debette and Markus, 2010; Jiang et al., 2018; Alber et al., 2019). According to the Fazekas visual rating scales, WMH can be classified into four grades (grade 0, no WMH; grade 1, focal or punctate lesions; grade 2, beginning confluent lesions; and grade 3, confluent lesions) (Prins and Scheltens, 2015). Severe WMH is also closely related to incidence rates of stroke, dementia, and death (Debette and Markus, 2010; Prins and Scheltens, 2015). However, the underlying mechanisms by which increasing WMH burden triggers cognitive impairment and vascular events remain largely unknown. A better understanding of the mechanisms linking WMH with cognitive decline is of utmost importance.

Resting-state functional MRI (rs-fMRI) is a non-invasive imaging technique used to explore the aberrant intrinsic functional architecture of the brain. It can be used to calculate the temporal correlations among intrinsic fluctuations of blood-oxygen-level-dependent (BOLD) signals across functionally related areas, also known as functional connectivity (FC) (Damoiseaux et al., 2006). Additionally, rs-fMRI also offers a good option to obtain a quantitative mapping of cerebrovascular reactivity (CVR) (Golestani et al., 2016b; Taneja et al., 2019), which has been applied in several diseases, including carotid artery stenosis/occlusion, Moyamoya disease (De Vis et al., 2018; Ryan et al., 2018; Taneja et al., 2019), cerebral small vessel disease (CSVD) (Ni et al., 2021), and hypertension (Ma et al., 2021). CVR is the change of cerebral blood flow in response to a vasodilatory challenge and reflects underlying vascular dysfunction or pathology (Sobczyk et al., 2014). Impaired CVR is closely related to normal aging (Lu et al., 2011) and might be a sensitive predictor of dementia (Buratti et al., 2015). However, the underlying mechanisms by which a decrease in CVR triggers cognitive decline remain largely unknown. Additionally, previous studies have indicated that the altered FC and CVR in WMH individuals were associated with cognitive decline (Chen et al., 2019; Ye et al., 2019; Ni et al., 2021; Vipin et al., 2021). Nonetheless, none of these studies investigated the interrelationships among CVR and FC in WMH individuals, and their combined influence on WMH-related cognitive decline has never been examined. A prior study suggested that a change of CVR is significantly associated with FC in motor and executive control networks in healthy adults (Golestani et al., 2016a). Hence, in the present study, we aim to explore the effects of CVR and FC on WMH-related cognitive decline and the exact relationship among WMH, CVR, FC, and cognitive decline. We applied rs-fMRI to evaluate CVR maps and FC in WMH-I, WMH-II, and WMH-III populations. We investigated the relationships among WMH burdens, impaired CVR, FC, and cognitive decline in WMH individuals. We hypothesized that severe WMH may lead to local CVR decline, then induces abnormal FC, and finally results in cognitive decline.

Abbreviations: BOLD, blood-oxygen-level-dependent; CI, confidence interval; CSVD, cerebral small vessel diseases; CVR, cerebrovascular reactivity; DWMH, deep white matter hyperintensity; FC, functional connectivity; FLAIR, fluid-attenuated inversion recovery; GRF, Gaussian random field; IPL, inferior parietal lobule; MMSE, Mini-Mental State Examination; MRI, magnetic resonance imaging; MOG, middle occipital gyrus; MoCA, Montreal Cognitive Assessment; PWMH, periventricular white matter hyperintensity; rs-fMRI, resting-state functional magnetic resonance imaging; SCWT, Stroop Color and Word Tests; TMT, Trail Making Test; WMH, white matter hyperintensity.

MATERIALS AND METHODS

Participants

The present research was conducted under the latest version of the Declaration of Helsinki and approved by the Nanjing Drum Tower Hospital Research Ethics Committee. All participants provided informed written consent before participating in the study. A total of 233 participants were consecutively recruited in Nanjing Drum Tower Hospital from January 2017 to April 2019. According to the Fazekas rating scale (Prins and Scheltens, 2015), the participants were divided into WMH-I (focal or punctate lesions, $n = 106$), WMH-II (beginning confluent lesions, $n = 72$), and WMH-III (confluent lesions, $n = 55$). WMH was categorized independently and unanimously by two radiologists, who visually evaluated the MRI without knowledge of the clinical profiles of the participants.

The inclusion criteria for WMH subjects were as follows: (Alber et al., 2019) age between 45 and 85 years old, Longstreth et al. (1996) the presence of mild to severe WMH on FLAIR, and (Han et al., 2018) no recent small subcortical infarctions (infarctions that presented a high signal in diffusion-weighted imaging were designated as a recent event). The exclusion criteria included the following: (Alber et al., 2019) a history of ischemic stroke with an infarct size larger than 1.5 cm in diameter or cardiogenic cerebral embolism, Longstreth et al. (1996) cerebral hemorrhage, Han et al. (2018) internal carotid artery or vertebral artery stenosis (>50%), Debette and Markus (2010) WMH due to immune-mediated demyelinating disease (multiple sclerosis, neuromyelitis optical, and acute disseminated encephalomyelitis), metabolic leukodystrophy, and genetic leukoencephalopathy, Jiang et al. (2018) cognitive impairments due to other neurological disorders, such as Alzheimer's disease, Parkinson's disease, and epilepsy (Prins and Scheltens, 2015) systemic diseases, such as cancer, shock, and anemia, Damoiseaux et al. (2006) prominent impairments of hearing or vision, and (Golestani et al., 2016b) MRI contraindications.

Neuropsychological Examinations

The cognitive evaluations were completed by two professional neuropsychologists. The obtained raw scores were firstly converted to *Z*-scores. The Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA)

were used to evaluate the global cognitive function. Other neuropsychological battery tests include Trail Making Tests A and B (TMT-A and TMT-B) and Stroop Color and Word Tests (SCWT) A, B, and C (SCWT-A, SCWT-B, and SCWT-C). In addition, both the Hamilton Depression Rating Scale and the Hamilton Anxiety Rating Scale were used to test the mental status of all subjects.

Magnetic Resonance Imaging Scanning

The rs-fMRI, 3D-T1 weighted, and 3D-FLAIR images were extracted from all subjects using a 3.0-T magnetic resonance scanner (Philips Medical Systems, Netherlands). The imaging parameters used were as follows: high-resolution T1-weighted turbo gradient echo sequence: repetition time (TR) = 9.8 ms, echo time (TE) = 4.6 ms, flip angle (FA) = 8°, field of view (FOV) = 250 × 250 mm^2, number of slices = 192, matrix sizes = 192 × 256 × 256, thickness = 1.0 mm; 3D-FLAIR images: TR = 4,500 ms, TE = 333 ms, FOV = 258 × 247 mm^2, number of slices = 200, matrix sizes = 200 × 272 × 261; thickness = 1.0 mm. The rs-fMRI was acquired by a gradient-echo-planar imaging sequence: TR = 2,000 ms, TE = 30 ms, FA = 90°, FOV = 192 × 192 mm^2, matrix size = 64 × 64, thickness = 4.0 mm, number of slices = 35; each functional image contained 240 volumes. During the scanning of rs-fMRI, the participants were instructed to keep their eyes closed, to stay as still as possible, and not to fall asleep or think about anything in particular. Additionally, T2-weighted and diffusion-weighted imaging sequences were collected to assess acute or sub-acute infarctions.

Volume Quantification of Brain and White Matter Hyperintensity

The volumes of total brain, gray matter, and white matter were automatically obtained using Statistical Parametric Mapping (SPM8[1]) based on 3D-T1 images. The Wisconsin White Matter Hyperintensity Segmentation Toolbox[2] was used to semiautomatically quantify total white matter hyperintensity (TWMH) volume, including deep WMH (DWMH) and periventricular WMH (PWMH) based on T1-weighted and FLAIR images.

Evaluation of Cerebrovascular Reactivity

The CVR maps based on BOLD signals of rs-fMRI was calculated through SPM12 and in-house MATLAB (MathWorks, Natick, MA, United States) scripts. The preprocessing procedures included the following: (Alber et al., 2019) head motion correction, Longstreth et al. (1996) spatial smoothing through convolution with an isotropic Gaussian kernel of 8 mm, and (Han et al., 2018) linear detrending. Given that the global BOLD fluctuations in the frequency range of 0.02–0.04 Hz mostly contributed to natural variation in end-tidal (Et) CO_2 levels (Liu et al., 2017), the rs-fMRI data were temporally filtered with a band-pass filter (filtering range: from 0.02 to 0.04 Hz). The average whole-brain rs-BOLD signal was calculated as the reference time course. A general linear model was employed in a voxel-wise manner, with the reference time course as the independent variable and the time course of each voxel as the dependent variable, generating a CVR index map. Then, the CVR index map was further normalized to the reference time course, yielding a relative CVR map. The relative CVR maps were spatially normalized to the standard Montreal Neurologic Institute (MNI) template, with a resampled voxel size of 3 × 3 × 3 mm^3 so that a global voxel-wise between-subject comparison could be conducted.

Functional Connectivity Analysis

The rs-fMRI was preprocessed using a toolbox for Data Processing and Analysis for Brain Imaging (DPABI) V2.3. The detailed preprocessing procedures were described in our previous study (Chen et al., 2019). Brain regions showing altered CVR among the three groups served as the seed to construct the resting-state FCs. Pearson cross-correlation analysis was applied between the seed time course and the time course of the whole brain voxels, and then Fisher's z-transformation was used to improve the normality of the correlation coefficients $z = 0.5 \times \ln[(1 + r)/(1 - r)]$. Ultimately, the individual FC maps of each region showing a significant group difference of CVR were obtained.

Statistical Analysis

White matter hyperintensity volumes were log-transformed to normalize the distribution. Differences in characteristics across the three groups were analyzed using one-way analysis of variance, chi-square (χ^2) test, and Kruskal–Wallis test by SPSS 22.0 software (IBM Corp., Armonk, NY, United States). $P < 0.05$ was considered statistically significant.

A voxel-wise one-way analysis of covariance was performed on the comparison of CVR and FC maps among three groups using DPABI V2.3, adjusting for age, gender, and education years. The threshold was set at $P < 0.001$ and corrected by Gaussian random field (GRF) correction (voxel p threshold for the minimum cluster size < 0.001 and cluster p threshold < 0.001) for multiple comparisons. The number of minimum cluster voxels was 100. Subsequently, for the GRF-corrected statistically significant brain regions, a *post hoc* analysis was performed to investigate group differences between two arbitrary groups, additionally correcting with Bonferroni correction for multiple comparisons ($P < 0.05/3$ was considered statistically significant). Then, partial correlation analysis was conducted to assess the relationship of altered CVR and FC with clinical neuropsychological variables after adjusting for variables known to affect cognitive function (i.e., age, gender, and years of education).

Additionally, serial mediation analysis was performed to explore the mediating effect of altered CVR and FC on the relationship between WMH and cognitive scores after adjusting for age, gender, education years. The indirect effect of WMH on cognitive scores was estimated by multiplying the coefficients from three sequential linear regression models representing the pathway links ($a \times b \times c$). The models simultaneously estimate two shorter pathways, from WMH to altered CVR to cognitive scores ($a \times b_2$) and from WMH to abnormal FC to cognition

[1]http://www.fil.ion.ucl.ac.uk/spm

[2]https://sourceforge.net/projects/w2mhs

scores ($a_2 \times c$), and the direct effect between WMH and cognitive scores adjusted for the indirect pathways ($d' = d - a \times b \times c - a \times b_2 - a_2 \times c$). We computed path coefficients and bias-corrected 95% confidence intervals for the size of the mediating effects with bootstrapping (k = 5,000 samples). The serial mediating effect is thought to be present if the 95% confidence interval (CI) does not contain zero. All the mediation analyses were conducted in PROCESS for the SPSS 22.0 framework.

RESULTS

Demographic, Clinical, and Cognitive Data Among the Three Groups

The basic characteristics of the three groups are presented in **Table 1**. Among the three groups, there were no significant differences in all of the demographic and clinical factors (gender, education years, vascular risk factors, mental status, total brain volume, gray, and white matter volume, and bilateral hypothalamus volume; all $P > 0.05$) except for age and WMH volume ($P < 0.05$). We thus removed the effect of age in the following analyses. With the Fazekas score increased, the volume of WMH increased accordingly ($P < 0.05$). Compared with the WMH-I subjects, the WMH-III and WMH-II subjects exhibited poorer MoCA, SCWT-A, SCWT-B, and TMT-B (all $P < 0.05$). The WMH-III subjects displayed poorer MMSE, SCWT-C, and TMT-A in contrast with the WMH-I subjects (all $P < 0.05$).

Cerebrovascular Reactivity Differences Among the Three Groups

Atlas maps of CVR based on the BOLD signal for each group are shown in **Figure 1A**. As shown in **Table 2** and **Figure 1B**, the three groups displayed a significantly different CVR in the left middle occipital gyrus (MOG) after adjusting for age, gender, and education years with GRF correction for multiple comparisons. A *post hoc* analysis showed that there was a decreasing trend of CVR in the left MOG, accompanied by the increasing WMH grades. The WMH-II and WMH-III groups displayed a lower CVR in the left MOG than the WMH-I group ($P = 0.008$ in the WMH-II group and $P < 0.001$ in the WMH-III group vs. the WMH-I group, Bonferroni-corrected, **Figure 1C**). Relative to the WMH-II group, the WMH-III group showed a lower CVR in the left MOG. However, it did not achieve statistical significance.

Functional Connectivity Differences Among the Three Groups

Since the three WMH groups displayed a significant difference of CVR in the left MOG, we further investigated the differences of FC pattern among the three groups when using the left MOG as seed. As shown in **Figure 2**, the three groups showed a significantly different FC of the left insula, left inferior parietal lobule (IPL), and thalamus with the left MOG, after adjusting for age, gender, and education years with GRF correction for multiple comparisons. The detailed coordinate information of the above-mentioned regions is shown in **Table 3**. *Post hoc* analysis revealed

TABLE 1 | Basic characteristics and cognitive evaluations among the three groups.

Items	WMH-I (n = 106)	WMH-II (n = 72)	WMH-III (n = 55)	F/χ^2	P-value
Age (years)	64.34 ± 6.85	66.78 ± 7.96	68.11 ± 6.47[a]	5.709	0.004*
Gender, male (% male)	55 (51.89)	37 (51.39)	24 (43.64)	1.093	0.579
Education (years)	11.42 ± 3.85	10.46 ± 3.87	10.24 ± 3.87	2.194	0.114
Hypertension, *n* (%)	65 (61.32)	53 (73.61)	37 (67.27)	2.926	0.231
Diabetes, *n* (%)	27 (25.47)	16 (22.22)	13 (23.63)	0.254	0.881
Smoking, *n* (%)	17 (16.04)	18 (25.00)	14 (25.45)	2.922	0.232
Hyperlipidemia, *n* (%)	20 (17.92)	20 (22.22)	11 (20.00)	2.141	0.343
WMH volume (ml)	1.94 ± 1.13	5.34 ± 1.86[a]	17.45 ± 10.44[ab]	93.88	<0.001*
Total brain volume (ml)	1,405.83 ± 116.18	1,410.40 ± 133.41	1,409.35 ± 135.44	0.03	0.974
Gray matter volume (ml)	572.42 ± 47.36	567.08 ± 50.51	551.45 ± 55.91	2.84	0.061
White matter volume (ml)	473.28 ± 46.79	477.23 ± 59.97	477.07 ± 58.24	0.13	0.88
HAMD	5.78 ± 4.62	6.00 ± 5.21	6.40 ± 4.92	0.270	0.764
HAMA	7.79 ± 6.51	9.32 ± 7.60	9.04 ± 7.50	1.141	0.321
***Z*-score of cognition**					
MMSE	0.20 ± 0.56	−0.09 ± 1.10	−0.27 ± 1.38[a]	4.418	0.013*
MoCA	0.45 ± 0.70	−0.26 ± 0.97[a]	−0.52 ± 1.19[a]	24.39	<0.001*
SCWT-A	−0.35 ± 0.47	0.11 ± 1.12[a]	0.51 ± 1.29[a]	14.26	<0.001*
SCWT-B	−0.25 ± 0.56	0.18 ± 1.41[a]	0.25 ± 0.96[a]	5.81	0.004*
SCWT-C	−0.24 ± 0.67	−0.08 ± 1.15	−0.35 ± 1.21[a]	6.08	0.003*
TMT-A	−0.23 ± 0.48	0.05 ± 1.15	0.38 ± 1.37[a]	6.83	0.001*
TMT-B	0.27 ± 0.97	0.14 ± 0.98[a]	0.35 ± 0.96[a]	6.71	0.002*

*Values are presented as means ± standard deviation, median with minimum, and maximum or absolute numbers with percentages. HAMA, Hamilton Anxiety Rating Scale; HAMD, Hamilton Depression Rating Scale; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; SCWT, Stroop Color and Word Tests; TMT, Trail Making Test; WMH, white matter hyperintensity. *$P < 0.05$, statistical difference among the three groups. [a]$P < 0.05$, statistically different from WMH-I. [b]$P < 0.05$, statistically different from WMH-II.*

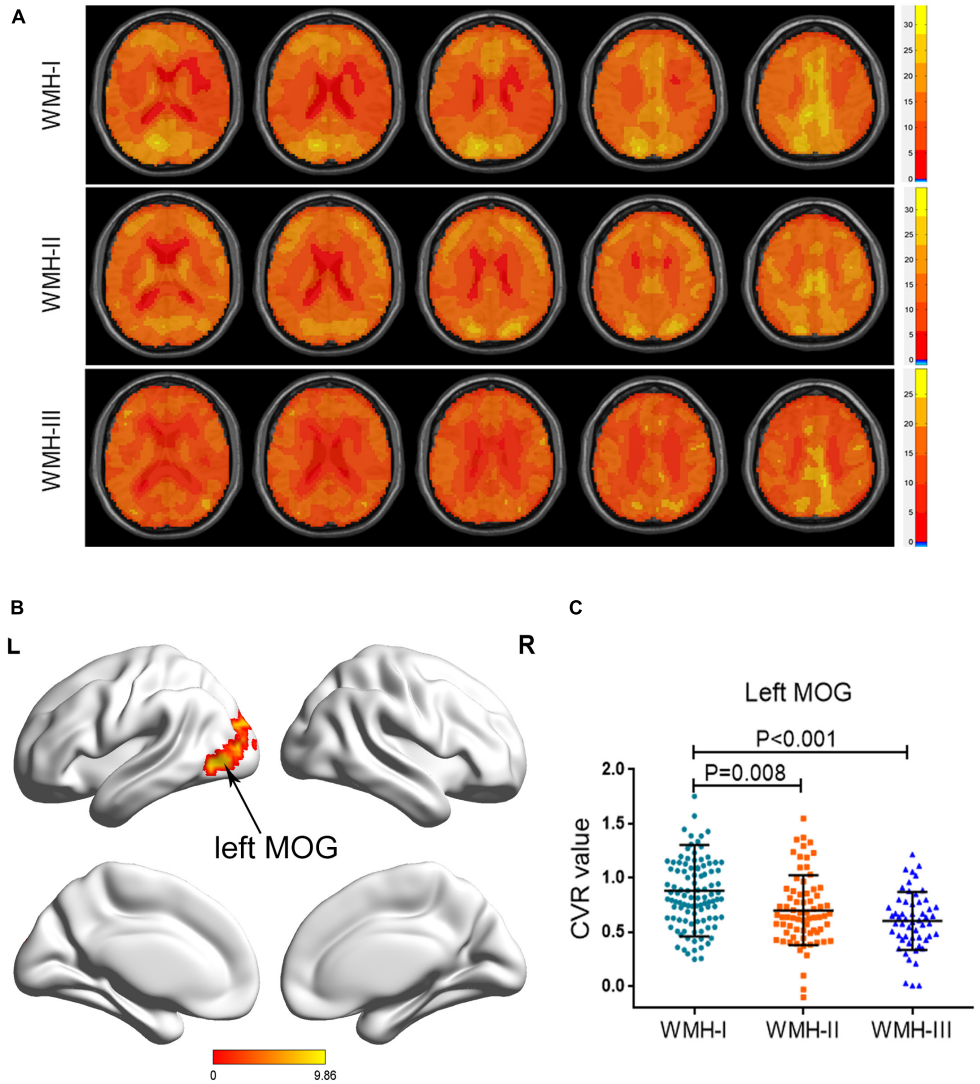


FIGURE 1 | Group differences of CVR across the three groups. **(A)** Axial CVR mapping for each WMH cohort. **(B)** The brain areas displayed a significant difference among the three groups (left MOG). **(C)** The subjects in the WMH-II and WMH-III groups displayed a significantly lower CVR in the left MOG than the subjects in the WMH-I group. CVR, cerebrovascular reactivity; MOG, middle occipital gyrus; WMH, white matter hyperintensity.

that subjects in WMH-III displayed a significantly greater FC of the left IPL-MOG and the left insula-MOG than subjects in the other two groups (**Figures 2B,C**). In terms of FC of the thalamus-MOG, subjects in the WMH-III and WMH-II groups displayed a significantly greater FC in the thalamus than the subjects in the WMH-I group (**Figure 2D**).

TABLE 2 | Brain region that displayed a different cerebrovascular reactivity among the three groups.

Brain areas	Peak MNI coordinates *x*, *y*, *z* (mm)	Peak *F* value	Number of cluster voxels
Left MOG	−30, −60, 21	10.488	285

The voxel Z threshold < 0.001 with Gaussian random field correction (voxel p threshold for the minimum cluster size < 0.001 and cluster p threshold < 0.001) for multiple comparisons. MNI, Montreal Neurological Institute; MOG, middle occipital gyrus.

Correlations Among Altered Cerebrovascular Reactivity, Functional Connectivity, and Cognitive Performance

We performed a partial correlation analysis to explore the correlations among altered CVR, FC, and cognitive performance after adjusting for age, gender, and education years in the WMH-III group. As shown in **Figure 3**, the lower CVR in the left MOG was associated with a poor MoCA score (r = 0.291, P = 0.036, **Figure 3A**) as well as a longer SCWT-A (r = −0.331, P = 0.016, **Figure 3B**), SCWT-C (r = −0.382, P = 0.005, **Figure 3C**), and

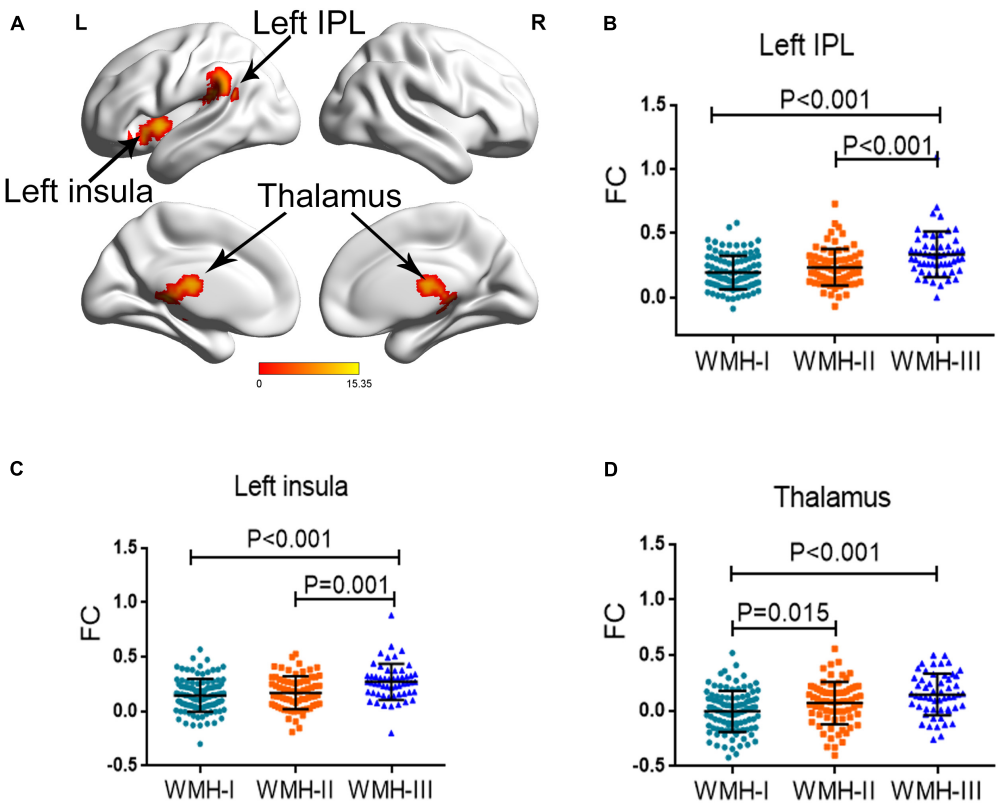


FIGURE 2 | Group differences of FC across the three groups. **(A)** The regions displayed a significant difference among the three groups (left insula, left IPL, and thalamus). FC, functional connectivity; IPL, inferior parietal lobule. **(B)** The subjects in WMH-III displayed a significantly greater FC in the left IPL than the subjects in the other two groups. **(C)** The subjects in the WMH-III group displayed a significantly higher FC in the left insula than the subjects in the other two groups. **(D)** The subjects in the WMH-III and WMH-II groups displayed a significantly greater FC in the thalamus than the subjects in the WMH-I group. FC, functional connectivity; IPL, inferior parietal lobule; WMH, white matter hyperintensity.

TMT-A time (r = −0.318, P = 0.005, **Figure 3D**). As shown in **Figure 4**, the FC of the left insula-MOG was positively associated with the MMSE score (r = 0.356, P = 0.010, **Figure 4A**) and the MoCA score (r = 0.401, P = 0.003, **Figure 4B**). The higher FC of the left insula-MOG was associated with a shorter SWCT-A (r = −0.308, P = 0.026, **Figure 4C**) and TMT-A time (r = −0.497, P < 0.001, **Figure 4D**). Similarly, the higher FC of the left IPL-MOG was associated with a shorter TMT-A time (r = −0.497, P = 0.018, **Figure 4E**). Nonetheless, there was no significant correlation of FC in the thalamus with cognitive performance. When analyzed the relationship between CVR and FC in the WMH-III group, we only found that the FC of the left insula was positively correlated with the CVR in the left MOG (r = 0.347, p = 0.012; the details are shown in **Supplementary Table 1**).

Furthermore, we performed a partial correlation analysis in the WMH-I and WMH-II groups, respectively. However, we just found that the CVR of the left MOG is negatively associated with the time of SCWT-C (r = −0.256, p = 0.008) in the WMH-I group (the detailed negative results of the partial correlation analysis in the WMH-I and WMH-II groups are shown in **Supplementary Tables 1–3**).

TABLE 3 | Brain areas that displayed different functional connectivity values among the three groups.

Brain areas	Peak MNI coordinates *x*, *y*, *z* (mm)	Peak *F* value	Number of cluster voxels
Left insula	−36, 6, 0	12.4497	116
Left IPL	−39, −39, 21	15.3541	252
Thalamus	−3, −18, 15	11.9213	252

The voxel Z threshold < 0.001 with Gaussian random field correction (voxel p threshold for the minimum cluster size < 0.001 and cluster p threshold < 0.001) for multiple comparisons. MNI, Montreal Neurological Institute; IPL, inferior parietal lobule.

Serial Mediating Effect of Abnormal Cerebrovascular Reactivity and Functional Connectivity on the Relationship Between White Matter Hyperintensity and Cognitive Performance

We performed serial mediation analysis to explore the mediating effect of altered CVR and FC on the relationships between WMH

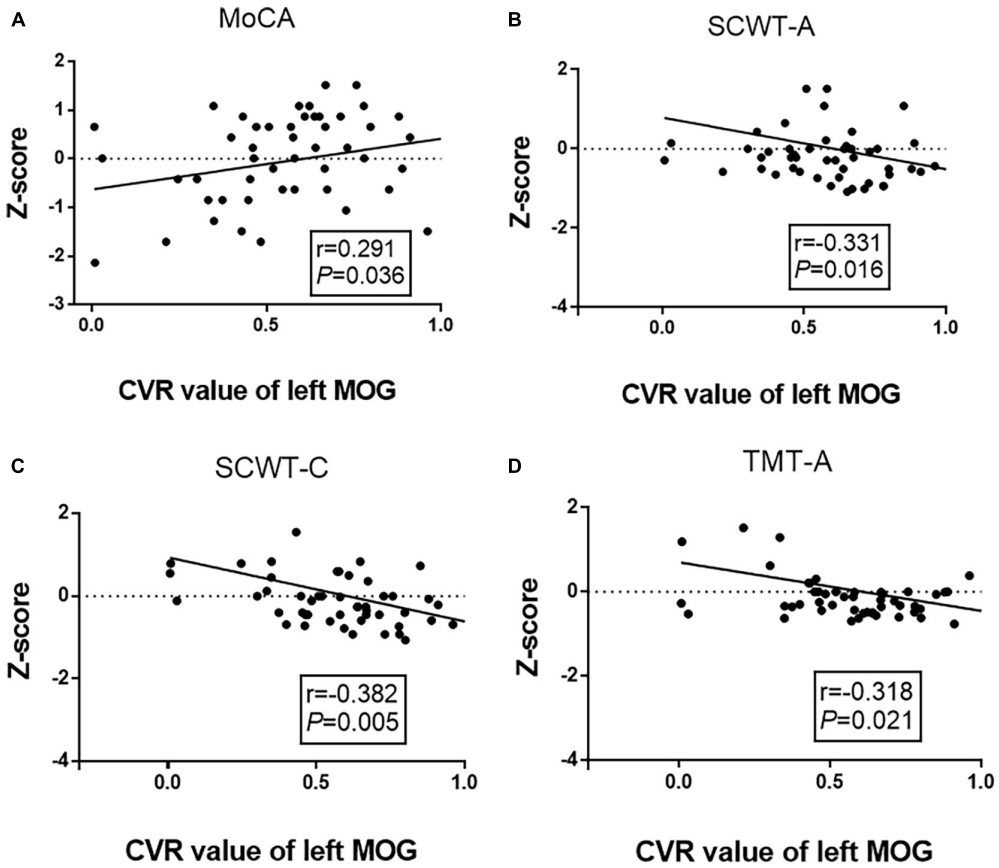


FIGURE 3 | Correlation analyses between cognitive function and CVR of the left MOG in the white matter hyperintensity (WMH)-III group. **(A)** The CVR of the left MOG was positively correlated with MoCA score (r = 0.291, P = 0.036). **(B)** The CVR of the left MOG was negatively correlated with the time of SCWT-A (r = –0.331, P = 0.016). **(C)** The CVR of the left MOG was negatively correlated with the time of SCWT-C (r = –0.382, P = 0.005). **(D)** The CVR of the left MOG was negatively correlated with the time of TMT-A (r = –0.318, P = 0.021). The scores were z-transformed scores of the relevant tests. CVR, cerebrovascular reactivity; MOG, middle occipital gyrus; MoCA, Montreal Cognitive Assessment; SCWT, Stroop Color and Word Tests; TMT, Trail Making Test.

and cognitive scores after adjusting for age, gender, and education years in the WMH-III group. We first found that the TWMH volume was indirectly associated with poorer MMSE score (indirect effect: −0.1499, 95% CI: −0.6165, −0.0213; **Figure 5A**) and MoCA scores (indirect effect: −0.1714, 95% CI: −0.5833, −0.0300; **Figure 5B**) and longer TMT-A time (indirect effect: β = 0.2216, 95% CI: 0.0056, 1.0234; **Figure 5C**), which was mediated serially by the aberrant CVR of the left MOG and FC of the left insula-MOG. When substituting the TWMH volumes with PWMH volumes or DWMH volumes in the mediation modal, we found that just the pathway of PWMH achieved significance, suggesting that the indirect effect of WMH on the above-mentioned cognitive scores was dominated by PWMH but not DWMH (the details are shown in **Figures 5D–F**).

Additionally, we also performed a *post hoc* mediation modal which hypothesized the WMH results from the CVR decline and then led to the FC change and cognitive decline in the WMH-III group. We did not find any mediation effect in this pathway (all 95% CI contained zero, **Supplementary Table 4**), suggesting that the mediation effect is specific to the WMH–CVR–FC–cognitive performance order.

DISCUSSION

In the present study, we demonstrated that, with the degree of WMH aggravated, the CVR of the left MOG decreased, and the FC of the left insula-MOG, left IPL-MOG, and thalamus-MOG abnormally changed. Notably, at the phase of severe WMH load (WMH-III), the aberrant CVR of the left MOG and FCs of the left insula/left IPL with MOG was associated with cognitive performance. Furthermore, PWMH contributes indirectly to poorer MMSE, MoCA, and TMT-A scores, specifically due to a lower CVR of the left MOG and a higher FC of the left insula-MOG, suggesting that severe WMH may lead to CVR decline, then induce abnormal FC, and finally result in cognitive decline. Additionally, this study provided a comprehensive understanding on the pathophysiology mechanisms of WMH progression and WMH-related cognitive decline.

The CVR patterns in WMH patients have been investigated by studies using transcranial doppler ultrasound or fMRI. Traditionally, the breath holding index of middle cerebral arteries *via* transcranial doppler ultrasound is useful for measuring global CVR, and previous researchers found that the global

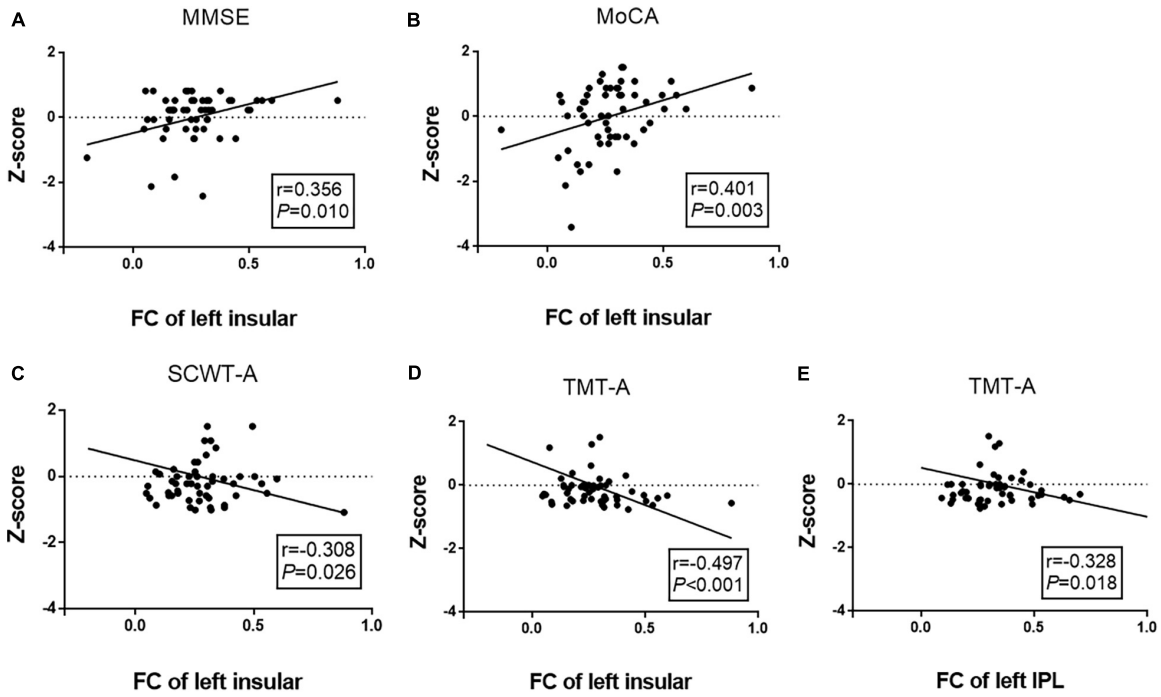


FIGURE 4 | Correlation analyses between FCs and cognitive function in the white matter hyperintensity (WMH)-III group. **(A)** The FC of the left insula-MOG was positively correlated with MMSE score (r = 0.356, P = 0.010). **(B)** The FC of the left insula-MOG was positively correlated with the MoCA score (r = –0.401, P = 0.003). **(C)** The FC of the left insula-MOG was negatively correlated with the time of SCWT-A (r = –0.308, P = 0.026). **(D)** The FC of the left insula-MOG was negatively correlated with the time of TMT-A (r = –0.497, P < 0.001). **(E)** The FC of the left IPL-MOG was negatively correlated with the time of TMT-A (r = –0.328, P = 0.018). The scores were z-transformed scores of the relevant tests. FC, functional connectivity; IPL, inferior parietal lobule; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; SCWT, Stroop Color and Word Tests; TMT, Trail Making Test.

CVR decreased as the WMH aggravated (Bakker et al., 1999; Fu et al., 2006). Resting fMRI technique has been used to detect the regional CVR in CSVD (Ni et al., 2021) and hypertensive patients (Ma et al., 2021). Decreased CVR has been shown across different cortical regions, including the frontal cortex, parietal cortex, and occipital cortex in WMH patient (Ni et al., 2021). In the present study, we found that moderate–severe-WMH individuals displayed a lower CVR in the left MOG than mild-WMH individuals, and the decreased CVR was associated with general cognition (MoCA), information processing speed (SCWT-A and TMT-A), and executive function (SCWT-C) in severe-WMH individuals. The occipital lobe contains most of the anatomical region of the visual cortex and contributes to visual information processing and communication with the cerebral cortex as well as plays a crucial role in the perception of facial emotion. A previous study showed that the amplitude of occipital sources of resting-state alpha rhythms is related to Alzheimer's disease neurodegeneration in the occipital lobe along with pathological aging (Babiloni et al., 2015). MOG plays an important role in category-selective attention modulating unconscious face/tool processing (Tu et al., 2013). Consequently, keeping in mind the above-mentioned overview, our finding may suggest that processing bias may be initiated as a perceptual visual bias, which may cause a series of cognitive decline.

Additionally, the FC patterns of the left insula-MOG, left IPL-MOG, and thalamus-MOG were abnormally increased when using the left MOG as seed. Increased FC might reflect an impaired balance between synaptic excitation and inhibition and the loss of the ability to inhibit activities. Furthermore, we found that the FC of the left insula-MOG was positively associated with general cognition (MMSE and MoCA) and information processing speed (SCWT-A and TMT-A). The FC of the left IPL-MOG was associated with information processing speed (TMT-A). The insula is instrumental in integrating disparate functional systems that are involved in cognition processing (Chang et al., 2013). It was reported that the function of the insula could be activated in a broad range of cognitive domains (Augustine, 1996; Shelley and Trimble, 2004). IPL is an important component of default mode network, which is known to be crucial in maintaining normal cognition (Buckner et al., 2008; Harrison et al., 2008). Consistent with these studies, our findings proved that the FC changes of the left insula and IPL were involved in WMH-related cognitive decline.

The causal relationships among WMH, CVR, FC, and cognitive functions remain unknown. We further explored this problem by multiple-step mediation analysis and confirm that severe WMH led to CVR decline of the left MOG, which then induced an abnormally higher FC between the left insula and MOG and finally resulted in cognitive decline.

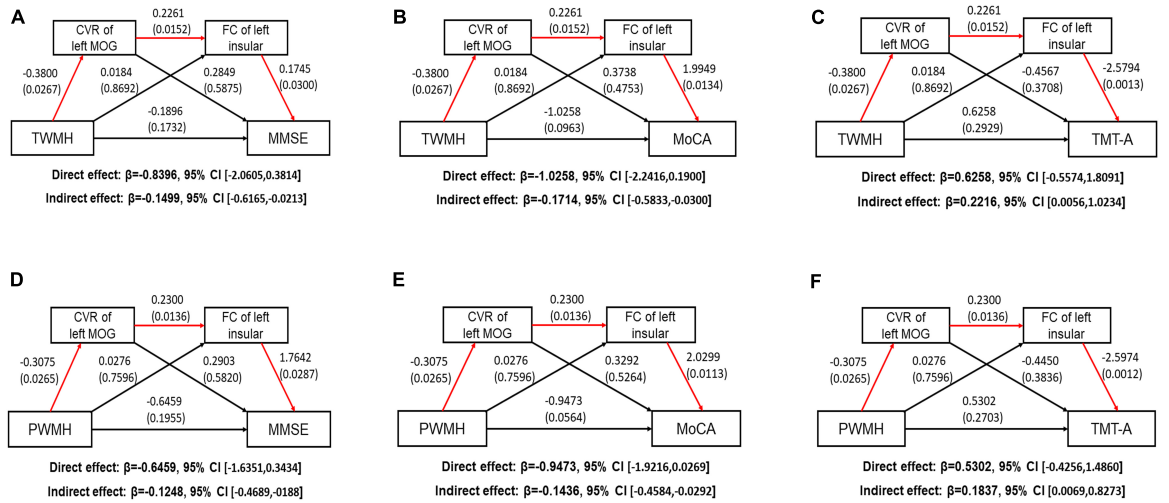


FIGURE 5 | Path models show the direct and indirect relationships between white matter hyperintensity volume and cognitive function. The significant pathways are highlighted in red lines. For each connection, the standard coefficient and *P*-value are shown. The direct and indirect effect and their 95% CI are also shown. **(A)** TWMH volume was indirectly associated with poorer MMSE score through aberrant CVR of left MOG and FC of left insula-MOG (Indirect effect: −0.1499, 95% CI [−0.6165, −0.0213]. **(B)** TWMH volume was indirectly associated with poorer MoCA score through aberrant CVR of left MOG and FC of left insula-MOG (Indirect effect: −0.1714, 95% CI [−0.5833, −0.0300]. **(C)** TWMH volume was indirectly associated with longer TMT-A time through aberrant CVR of left MOG and FC of left insula-MOG (Indirect effect: −0.2218, 95% CI [0.0056, 1.0234]. **(D)** PWMH volume was indirectly associated with poorer MMSE score through aberrant CVR of left MOG and FC of left insula-MOG (Indirect effect: −0.1248, 95% CI [−0.4689, −0.0188]. **(E)** PWMH volume was indirectly associated with poorer MoCA score through aberrant CVR of left MOG and FC of left insula-MOG (Indirect effect: −0.1436, 95% CI [−0.4584, −0.0292]. **(F)** PWMH volume was indirectly associated with longer TMT-A time through aberrant CVR of left MOG and FC of left insula-MOG (Indirect effect: 0.1837, 95% CI [0.0069, 0.8273]. CI, confidence interval; CVR, cerebrovascular reactivity; FC, functional connectivity; MOG, middle occipital gyrus; TWMH, total white matter hyperintensity; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; TMT, Trail Making Test; PWMH, periventricular white matter hyperintensity.

Conventionally, hemodynamic dysfunction, represented by CVR, is an etiological factor of WMH, which has been demonstrated by some animal model (Shibata et al., 2004; Ihara and Tomimoto, 2011) and clinical research (Sam et al., 2016a,b). However, mediation of the present study first revealed that severe WMH may lead to CVR decline in the left MOG, and the pathway just achieved significance in PWMH but not in DWMH. A previous study suggested that WMH can disrupt the left inferior fronto-occipital fasciculus, which is one of the connections of the left MOG (Chen et al., 2020). PWMH is more likely to disrupt the white matter microstructure than DWMH (Griffanti et al., 2018; Yang et al., 2020). From the view of the neurovascular unit, corresponding neuron cell body injury will be triggered after a severe axon injury (WMH-III), and then peripheral microvascular will be injured secondary, leading to CVR changes. Consistent with previous studies, we confirmed that severe PWMH is more likely to induce structural or functional disorders and then lead to cognitive decline than DWMH. Furthermore, our findings also demonstrated that the left MOG may act as a connection, which is involved in the processing of cognitive decline by being connected with the insular–cortical regions in severe WMH individuals.

The current study is innovative and acceptable for the following aspects: Firstly, rs-fMRI does not require any specialized cooperation, resulting in higher patient compliance and research accuracy. Secondly, the WMH subjects that we recruited in this work were relatively homogeneous and restricted to CSVD populations. Additionally, in the partial correlation and mediation analysis, we adjusted for all possible potential confounders, including age, gender, and education years.

Several limitations of the present work should be addressed. Firstly, no healthy controls were included in our study. A previous study using transcranial doppler ultrasound indicated that there was no significant difference in CVR between healthy controls and WMH-I groups (Bian et al., 2019). Future works should recruit healthy controls to verify whether the CVR of WMH-I subjects is lower than that of healthy controls. Secondly, we did not consider the impact of regional gray matter atrophy on FC and cognitive decline, which will have some influence on the generalization of the results. Finally, the sample is relatively small, and the nature of this study is cross-sectional. No causal inferences or directionality can be made. We are continuing to follow them up to validate our findings.

CONCLUSION

The present study demonstrated that aberrant CVR decline in the left MOG and FC of the left insula with the left MOG attributed to WMH progression were responsible for the poor general cognition and information processing speed in severe-WMH individuals. Our findings may help to understand the mechanisms underlying the onset of cognitive decline in severe-WMH individuals and to explore novel brain markers of vascular cognitive impairment.

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 Nanjing Drum Tower Hospital Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YX designed this study. DY wrote the manuscript and statistically analyzed the data. RQ, LC, HX, and LN helped analyze MRI data. JM, PS, LH, and BZ helped collect data. YX and MZ revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was supported by the National Key Research and Development Program of China (2016YFC1300504), the National Natural Science Foundation of China (81630028 and 81971112), the Key Research and Development Program of Jiangsu Province of China (BE2020620), and Jiangsu Province Key Medical Discipline (ZDXKA2016020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.807585/full#supplementary-material

REFERENCES

Alber, J., Alladi, S., Bae, H. J., Barton, D. A., Beckett, L. A., Bell, J. M., et al. (2019). White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): knowledge gaps and opportunities. *Alzheimers Dement* 5, 107–117. doi: 10.1016/j.trci.2019.02.001

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res. Brain Res. Rev.* 22, 229–244. doi: 10.1016/s0165-0173(96)00011-2

Babiloni, C., Del Percio, C., Boccardi, M., Lizio, R., Lopez, S., Carducci, F., et al. (2015). Occipital sources of resting-state alpha rhythms are related to local gray matter density in subjects with amnesic mild cognitive impairment and Alzheimer's disease. *Neurobiol. Aging* 36, 556–570. doi: 10.1016/j.neurobiolaging.2014.09.011

Bakker, S. L., de Leeuw, F. E., de Groot, J. C., Hofman, A., Koudstaal, P. J., and Breteler, M. M. (1999). Cerebral vasomotor reactivity and cerebral white matter lesions in the elderly. *Neurology* 52, 578–583. doi: 10.1212/wnl.52.3.578

Bian, Y., Wang, J. C., Sun, F., Sun, Z. Y., Lin, Y. J., Liu, Y., et al. (2019). Assessment of cerebrovascular reserve impairment using the breath-holding index in patients with leukoaraiosis. *Neural Regener. Res.* 14, 1412–1418. doi: 10.4103/1673-5374.251332

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function and relevance to disease. *Annals N. Y. Acad. Sci.* 1124, 1–38.

Buratti, L., Balestrini, S., Altamura, C., Viticchi, G., Falsetti, L., Luzzi, S., et al. (2015). Markers for the risk of progression from mild cognitive impairment to Alzheimer's disease. *JAD* 45, 883–890. doi: 10.3233/jad-143135

Chang, L. J., Yarkoni, T., Khaw, M. W., and Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cereb Cortex* 23, 739–749. doi: 10.1093/cercor/bhs065

Chen, H. F., Huang, L. L., Li, H. Y., Qian, Y., Yang, D., Qing, Z., et al. (2020). Microstructural disruption of the right inferior fronto-occipital and inferior longitudinal fasciculus contributes to WMH-related cognitive impairment. *CNS Neurosci. Therap.* 26, 576–588. doi: 10.1111/cns.13283

Chen, X., Huang, L., Ye, Q., Yang, D., Qin, R., Luo, C., et al. (2019). Disrupted functional and structural connectivity within default mode network contribute to WMH-related cognitive impairment. *Neuroimage Clin.* 24:102088. doi: 10.1016/j.nicl.2019.102088

Damoiseaux, J. S., Rombouts, S. A., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., et al. (2006). Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13848–13853. doi: 10.1073/pnas.0601417103

De Vis, J. B., Bhogal, A. A., Hendrikse, J., Petersen, E. T., and Siero, J. C. W. (2018). Effect sizes of BOLD CVR, resting-state signal fluctuations and time delay measures for the assessment of hemodynamic impairment in carotid occlusion patients. *Neuroimage* 179, 530–539. doi: 10.1016/j.neuroimage.2018.06.017

Debette, S., and Markus, H. S. (2010). The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ* 341:c3666. doi: 10.1136/bmj.c3666

Fu, J. H., Lu, C. Z., Hong, Z., Dong, Q., Ding, D., and Wong, K. S. (2006). Relationship between cerebral vasomotor reactivity and white matter lesions in elderly subjects without large artery occlusive disease. *J. Neuroimag.* 16, 120–125. doi: 10.1111/j.1552-6569.2006.00030.x

Golestani, A. M., Kwinta, J. B., Strother, S. C., Khatamian, Y. B., and Chen, J. J. (2016a). The association between cerebrovascular reactivity and resting-state fMRI functional connectivity in healthy adults: the influence of basal carbon dioxide. *Neuroimage* 132, 301–313. doi: 10.1016/j.neuroimage.2016.02.051

Golestani, A. M., Wei, L. L., and Chen, J. J. (2016b). Quantitative mapping of cerebrovascular reactivity using resting-state BOLD fMRI: validation in healthy adults. *Neuroimage* 138, 147–163. doi: 10.1016/j.neuroimage.2016.05.025

Griffanti, L., Jenkinson, M., Suri, S., Zsoldos, E., Mahmood, A., Filippini, N., et al. (2018). Classification and characterization of periventricular and deep white matter hyperintensities on MRI: a study in older adults. *Neuroimage* 170, 174–181. doi: 10.1016/j.neuroimage.2017.03.024

Han, F., Zhai, F. F., Wang, Q., Zhou, L. X., Ni, J., Yao, M., et al. (2018). Prevalence and Risk Factors of Cerebral Small Vessel Disease in a Chinese Population-Based Sample. *J. Stroke* 20, 239–246. doi: 10.5853/jos.2017.02110

Harrison, B. J., Pujol, J., López-Solà, M., Hernández-Ribas, R., Deus, J., Ortiz, H., et al. (2008). Consistency and functional specialization in the default mode brain network. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9781–9786. doi: 10.1073/pnas.0711791105

Ihara, M., and Tomimoto, H. (2011). Lessons from a mouse model characterizing features of vascular cognitive impairment with white matter changes. *J. Aging Res.* 2011:978761. doi: 10.4061/2011/978761

Jiang, J., Paradise, M., Liu, T., Armstrong, N. J., Zhu, W., Kochan, N. A., et al. (2018). The association of regional white matter lesions with cognition in a community-based cohort of older individuals. *Neuroimage Clin.* 19, 14–21. doi: 10.1016/j.nicl.2018.03.035

Liu, P., Li, Y., Pinho, M., Park, D. C., Welch, B. G., and Lu, H. (2017). Cerebrovascular reactivity mapping without gas challenges. *Neuroimage* 146, 320–326. doi: 10.1016/j.neuroimage.2016.11.054

Longstreth, W. T. Jr., Manolio, T. A., Arnold, A., Burke, G. L., Bryan, N., Jungreis, C. A., et al. (1996). Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301 elderly people. The Cardiovascular Health Study. *Stroke* 27, 1274–1282. doi: 10.1161/01.str.27.8.1274

Lu, H., Xu, F., Rodrigue, K. M., Kennedy, K. M., Cheng, Y., Flicker, B., et al. (2011). Alterations in cerebral metabolic rate and blood supply across the adult lifespan. *Cereb Cortex* 21, 1426–1434. doi: 10.1093/cercor/bhq224

Ma, J., Cao, X., Chen, F., Ye, Q., Qin, R., Cheng, Y., et al. (2021). Exosomal MicroRNAs Contribute to Cognitive Impairment in Hypertensive Patients by Decreasing Frontal Cerebrovascular Reactivity. *Front. Neurosci.* 15:614220. doi: 10.3389/fnins.2021.614220

Ni, L., Zhang, B., Yang, D., Qin, R., Xu, H., Ma, J., et al. (2021). Lower Cerebrovascular Reactivity Contributed to White Matter Hyperintensity-Related Cognitive Impairment: a Resting-State Functional MRI Study. *JMRI* 53, 703–711. doi: 10.1002/jmri.27376

Prins, N. D., and Scheltens, P. (2015). White matter hyperintensities, cognitive impairment and dementia: an update. *Nat. Rev. Neurol.* 11, 157–165. doi: 10.1038/nrneurol.2015.10

Ryan, C. M., Battisti-Charbonney, A., Sobczyk, O., Mikulis, D. J., Duffin, J., Fisher, J. A., et al. (2018). Evaluation of Cerebrovascular Reactivity in Subjects with and without Obstructive Sleep Apnea. *J. Stroke Cerebrovasc. Dis.* 27, 162–168. doi: 10.1016/j.jstrokecerebrovasdis.2017.08.015

Sam, K., Conklin, J., Holmes, K. R., Sobczyk, O., Poublanc, J., Crawley, A. P., et al. (2016a). Impaired dynamic cerebrovascular response to hypercapnia predicts development of white matter hyperintensities. *Neuroimage Clin.* 11, 796–801. doi: 10.1016/j.nicl.2016.05.008

Sam, K., Crawley, A. P., Poublanc, J., Conklin, J., Sobczyk, O., Mandell, D. M., et al. (2016b). Vascular Dysfunction in Leukoaraiosis. *AJNR Am. J. Neuroradiol.* 37, 2258–2264. doi: 10.3174/ajnr.A4888

Shelley, B. P., and Trimble, M. R. (2004). The insular lobe of Reil–its anatamico-functional, behavioural and neuropsychiatric attributes in humans–a review. *World J. Biol. Psychiatry* 5, 176–200. doi: 10.1080/15622970410029933

Shibata, M., Ohtani, R., Ihara, M., and Tomimoto, H. (2004). White matter lesions and glial activation in a novel mouse model of chronic cerebral hypoperfusion. *Stroke* 35, 2598–2603. doi: 10.1161/01.str.0000143725.19053.60

Sobczyk, O., Battisti-Charbonney, A., Fierstra, J., Mandell, D. M., Poublanc, J., Crawley, A. P., et al. (2014). A conceptual model for CO2-induced redistribution of cerebral blood flow with experimental confirmation using BOLD MRI. *Neuroimage* 92, 56–68. doi: 10.1016/j.neuroimage.2014.01.051

Taneja, K., Lu, H., Welch, B. G., Thomas, B. P., Pinho, M., Lin, D., et al. (2019). Evaluation of cerebrovascular reserve in patients with cerebrovascular diseases using resting-state MRI: a feasibility study. *Magnetic Reson. Imaging* 59, 46–52. doi: 10.1016/j.mri.2019.03.003

Tu, S., Qiu, J., Martens, U., and Zhang, Q. (2013). Category-selective attention modulates unconscious processes in the middle occipital gyrus. *Consciousness Cogn.* 22, 479–485. doi: 10.1016/j.concog.2013.02.007

Vipin, A., Wong, B., Kumar, D., Low, A., Ng, K., and Kandiah, N. (2021). Association between white matter hyperintensity load and grey matter atrophy in mild cognitive impairment is not unidirectional. *Aging* 13, 10973–10988. doi: 10.18632/aging.202977

Yang, D., Huang, L., Luo, C., Li, M., Qin, R., Ma, J., et al. (2020). Impaired Structural Network Properties Caused by White Matter Hyperintensity Related to Cognitive Decline. *Front. Neurol.* 11:250. doi: 10.3389/fneur.2020.00250

Ye, Q., Chen, X., Qin, R., Huang, L., Yang, D., Liu, R., et al. (2019). Enhanced Regional Homogeneity and Functional Connectivity in Subjects With White Matter Hyperintensities and Cognitive Impairment. *Front. Neurosci.* 13:695. doi: 10.3389/fnins.2019.00695

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, Qin, Chu, Xu, Ni, Ma, Shao, Huang, Zhang, Zhang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fnins.2021.739535

AT-NeuroEAE: A Joint Extraction Model of Events With Attributes for Research Sharing-Oriented Neuroimaging Provenance Construction

*Shaofu Lin[1,2], Zhe Xu[1], Ying Sheng[1], Lihong Chen[1,3] and Jianhui Chen[1,4,5]**

[1] Faculty of Information Technology, Beijing University of Technology, Beijing, China, [2] Beijing Institute of Smart City, Beijing University of Technology, Beijing, China, [3] Engineering Research Center of Digital Community, Beijing University of Technology, Beijing, China, [4] Beijing Key Laboratory of Magnetic Resonance Imaging (MRI) and Brain Informatics, Beijing University of Technology, Beijing, China, [5] Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing University of Technology, Beijing, China

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Karl Helmer, Massachusetts General Hospital and Harvard Medical School, United States
Yufeng Yao, Changshu Institute of Technology, China

***Correspondence:**
Jianhui Chen
chenjianhui@bjut.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *11 July 2021*
Accepted: *20 December 2021*
Published: *07 March 2022*

Citation:
Lin S, Xu Z, Sheng Y, Chen L and Chen J (2022) AT-NeuroEAE: A Joint Extraction Model of Events With Attributes for Research Sharing-Oriented Neuroimaging Provenance Construction. Front. Neurosci. 15:739535. doi: 10.3389/fnins.2021.739535

Provenances are a research focus of neuroimaging resources sharing. An amount of work has been done to construct high-quality neuroimaging provenances in a standardized and convenient way. However, besides existing processed-based provenance extraction methods, open research sharing in computational neuroscience still needs one way to extract provenance information from rapidly growing published resources. This paper proposes a literature mining-based approach for research sharing-oriented neuroimaging provenance construction. A group of neuroimaging event-containing attributes are defined to model the whole process of neuroimaging researches, and a joint extraction model based on deep adversarial learning, called AT-NeuroEAE, is proposed to realize the event extraction in a few-shot learning scenario. Finally, a group of experiments were performed on the real data set from the journal PLOS ONE. Experimental results show that the proposed method provides a practical approach to quickly collect research information for neuroimaging provenance construction oriented to open research sharing.

Keywords: neuroimaging provenance, neuroimaging text mining, event extraction, attribute extraction, deep adversarial learning

INTRODUCTION

Open and FAIR (Findable, Accessible, Interoperable, and Reusable) sharing (Poldrack and Gorgolewski, 2015; Abrams et al., 2021) of neuroimaging data has been widely recognized and has been a concern in computational neuroscience. From fMRIDC (Van Horn et al., 2001) and OpenfMRI (Poldrack and Gorgolewski, 2015) to OpenNeuro (Gorgolewski et al., 2017), a series of public neuroimaging databases have been constructed and have played an important role in brain and intelligence researches. Only taking the INDI database as an example, by March 22, 2017, Google Scholar search results showed that 913 publications clearly indicated the use of INDI data, with a total of 20,697 citations (Milham et al., 2018). Besides sharing raw experimental data,

research information sharing[1] has also become an important way of neuroimaging data sharing (Poline et al., 2012), which can effectively avoid various challenges of sharing raw experimental data, such as cost, privacy protection, and data ownership. BrainMap[2] collects the published information related to brain activated areas from the neuroimaging article to support coordinate based meta-analysis (Laird et al., 2005). Neurosynth[3] provides a specialized tool to automatically extract coordinates of activated brain area and research topic information from published neuroimaging articles and shared them by a unified platform (Yarkoni et al., 2011). NeuroVault[4] collected non-threshold statistical graphs and related information from different computational neuroscience to support graph based meta-analysis (Gorgolewski et al., 2016). These efforts on sharing neuroimaging research information enrich the content and method of neuroimaging data sharing and effectively support the brain and intelligence studies in computational neuroscience. For example, statistics show that the number of papers based on the neurosynth platform has reached 14,371 by July 2018. Related studies are involved with the brain mechanism of cognitive states (Alcalá-López et al., 2017), functional analysis of brain areas (Genon et al., 2018), large-scale brain structure-function mappings (Bolt et al., 2020), and so on.

Provenances (Moreau et al., 2008) are a research issue of neuroimaging data sharing Poldrack and Gorgolewski (2014) pointed out that sharing task-based fMRI (functional magnetic resonance imaging) data requires the appropriate descriptions of how they were acquired, including not only the MRI acquisition parameters but also the specific order and timing of stimulus presentation during the task, and how they have been transformed. These descriptions are provenances. The special topic "Collaborative efforts for understanding the human brain" of the journal *Frontiers in Neuroinformatics* emphasized that the key to controlling the re-executability of the publication of computational neuroscience is the generation and reporting, at all stages of the process, machine readable provenance documentation (Kennedy et al., 2019). From the data model (Keator et al., 2013; Maumet et al., 2016), the management tool (Keator et al., 2019) to the acquisition platform (Gorgolewski et al., 2016), an amount of work has been done to construct high-quality neuroimaging provenances in a standardized and convenient way. The neuroimaging provenances have been not only metadata of task-based experimental data but also structured descriptions of research cases in computational neuroscience, which collect massive research information about the whole research process of computational science to support multi-aspect resource integration, fast hypothesis generation, large-scale metadata analysis, strict result evaluation, etc., for realizing open and FAIR neuroscience (Abrams et al., 2021). However, existing studies on neuroimaging construction mainly adopt experts or process recording-based methods to obtain neuroimaging provenances. All methods are centered on the actual experimental data. Besides them, open research sharing in computational neuroscience still needs one new way to automatically and quickly extract provenance information from rapidly growing published sources.

Neuroimaging text mining recognizes important information from neuroimaging texts, especially neuroimaging articles, and provides a practical way to automatically extract provenance information from published sources. In recent years, related studies mainly focused on recognizing neuroimaging entities (Abacha et al., 2017; Shardlow et al., 2018; Riedel et al., 2019), hierarchical terminology system (Huangfu et al., 2020), and span of interests (Zhu et al., 2020). Because extracted information is mainly some words or phrases, and lacks rich relations, especially non-taxonomical relations, these studies can only obtain some fragmented key points of the research process and cannot provide a vivid process description. Hence, it is necessary to develop a new technology of neuroimaging text mining for research sharing-oriented neuroimaging provenance construction.

Based on this observation, this paper proposes a literature mining-based approach for neuroimaging provenance construction. The rest of this paper is organized as follows. Section "Related Work" summarizes previous work-related to neuroimaging provenances, neuroimaging text mining, and biomedical event extraction. Section "Contributions" introduces our contributions. Section "Materials and Methods" describes the proposed method. Experiments and results are presented in section "Results" to validate the effectiveness of the proposed method. Finally, section "Discussion" gives concluding remarks.

RELATED WORK

Neuroimaging Provenances

Provenances can be described in various terms depending on the domains where they are applied (Simmhan et al., 2005). Buneman et al. (2004) defined provenances as the description of the origins of data and the process by which it arrived at the database. Lanter (1991) characterized provenances as information describing materials and transformations applied to derive the data. Greenwood et al. (2003) viewed provenances as metadata recording the process of experiment workflows, annotations, and notes. Simmhan et al. (2005) defined provenances as information that helped determine the derivation history of a data product. Related studies have been widely concerned by many domains. Science magazine pointed out that provenance is an important element of service quality control (Foster, 2005). The NSF (National Science Foundation, Alexandria, VA, United States) Advisory Committee believed that a robust and stable cyberinfrastructure should support provenances (Task Force on Grand Challenges, 2011). The World Wide

[1]It is difficult to distinguish information from knowledge clearly because they have varying definitions in different applications. In this study, "research information" refers to all descriptions about computational science in scientific literature. It is involved with not only research findings, such as coordinates of activated brain areas, but also the whole research process, from experimental design to data analysis.

[2]http://www.brainmap.org/

[3]http://www.neurosynth.org/

[4]http://neurovault.org

Web Consortium standard of provenances, called PROV-DM (Provenance Data Model),[5] has been released in 2013. A series of international conferences about provenances, such as IPAW (International Provenance and Annotation Workshop, since 2006) and TaPP (Workshop on Theory and Practice of Provenance, since 2009), has attracted a large number of researchers.

Computational neuroscience is an important application domain of provenances. Neuroimaging provenances are the research focus. As early as 2006, the first fMRI data provenance challenge has been held (Moreau et al., 2008). A group of minimum provenance information (MI) guidelines (Gibson et al., 2008; Poldrack et al., 2008; Frishkoff et al., 2011) has been proposed for different types of neuroimaging data. Guided by these MI guidelines, the main neuroimaging data model, such as DICOM (Digital Imaging and Communications in Medicine)[6] and XCEDE (XML-based Clinical and Experimental Data Exchange),[7] provides the support for representing and storing provenance information. However, the provenance information storage embedded into experimental data models hampers the representation of more complex relationships among data (Ruiz-Olazar et al., 2016). Provenances independent from experimental data have drawn attentions. The representative research work is NIDM (Neuroimaging Data Model), which provides a complete description of provenance for computational neuroscience, from raw data to the final results including all the steps in between (Maumet et al., 2016). NIDM builds upon PROV-DM and extends it with terms found in XCEDE and DICOM headers. It consists of NIDM-Experiment, NIDM-Workflow, and NIDM-Results; NIDM-Experiment and NIDM-Workflow (Keator et al., 2013) model acquisition and processing of experimental data, respectively. NIDM-Results (Maumet et al., 2016) focuses on neuroimaging statistical results along with key image data summarizing the experiment.

In recent years, many studies are working to support the construction of neuroimaging provenances based on NIDM for realizing machine readable descriptions of the data collection, the processing workflow and environment, and the statistical procedures and results. The export and import of NIDM documents have been realized in many widely used neuroimaging tools and platforms, such as SPM, FSL, and NeuroVault. Keator et al. (2019) developed a python-based API (PyNIDM) which followed the simple organizational structure of NIDM-Experiment with functions to create, query, export, import, and transform NIDM-Experiment documents. Kennedy et al. (2019) developed the function to generate NIDM based provenances on ReproNim, which is a center for reproducible neuroimaging computation. Maumet et al. (2019) built a JSON-LD representation for NIDM-Results data and exposed it to the neuroimaging data management tool Datalad. With the growth of scale, the applications of neuroimaging provenances also have broken through traditional data descriptions oriented to data sharing (Van Horn et al., 2001) and process recordings oriented to analysis (Dinov et al., 2009), and are moved toward verification of scientific outputs (Arshad et al., 2019), research artifact sharing (Maumet, 2020), etc. Provenance-based research sharing (Yuan et al., 2018), which uses provenances as "research objects" (Belhajjame et al., 2015; Miksa and Rauber, 2017; That et al., 2017) (i.e., digital research artifacts for knowledge sharing and reproducibility in computational experiments), is becoming a trend for open and FAIR neuroscience.

However, provenance-based research sharing needs large-scale and high-quality neuroimaging provenances. Existing studies of neuroimaging provenances mainly adopt experts or process recordings-based construction methods. Public neuroimaging data or derived data sharing platforms, such as BrainMap and NeuroVault, rely on experts' manual annotations to collect neuroimaging provenance information. The obtained provenances vary dramatically in quality and grow slowly. Even if Brainspell[8] adopts the crowdsourcing-based annotation mode (Badhwar et al., 2016), the situation has not been fundamentally improved. Aiming at this problem, many neuroimaging tools, such as SPM, FSL, and PyNIDM, have provided the supporting for process-based provenance extraction. They can collect provenance information at the time of data acquisition or processing and do not require any extra effort. Process-based provenance extraction has become the most important means to construct neuroimaging provenances for large-scale data sharing. However, this kind of provenance extraction method is centered on actual experimental data. At present, neuroimaging data sharing still faces some important obstacles, such as privacy and data ownership. Public data account for only a small part of the total data. Similarly, many researchers do not actively generate and share neuroimaging provenances during the data acquisition or processing. Therefore, besides collecting provenance information during data acquisition or processing, it is also necessary to develop one new way to extract provenance information from published sources, especially published scientific articles.

Neuroimaging Text Mining

Neuroimaging technologies can non-invasively detect the connection between cognitive states and patterns of brain activity, and have received widespread attentions in computational neuroscience. Neuroimaging texts, especially scientific articles, are growing fast. Taking only fMRI (Wegrzyn et al., 2018) as an example, 307 relevant articles have been published in the journal PLOS ONE in 2020. These neuroimaging texts are valuable knowledge resources for studying human intelligence, pathological mechanism of brain and mental diseases, brain-computer interface, and so on. In the last decade, various neuroimaging meta-analysis (Yarkoni et al., 2011; Neumann et al., 2016) and collaborative analysis (Lei et al., 2020) have made great achievements. The value of neuroimaging texts as knowledge resources is already obvious. How to automatically and continuously extract knowledge from neuroimaging texts has become a key issue for open and FAIR neuroscience.

[5]http://www.w3.org/TR/prov-dm/

[6]http://medical.nema.org/dicom/

[7]https://github.com/incf-nidash/XCEDE

[8]http://brainspell.org

Neuroimaging text mining is an important branch of biomedical text mining and provides an effective approach to extract knowledge from neuroimaging texts. Early studies mainly focused on extracting neuroimaging knowledge for decoding a wide range of cognitive states Naud and Usui (2008) constructed the neuroscience terminology space by using the Vector Space Model and performed k-means based cluster on the abstract of neuroscience articles for extracting cognitive state-related topics. Neurosynth (Yarkoni et al., 2011) recognized terms based on word frequency and used the Naive Bayesian Classifier to predict the occurrence of specific terms based on the whole-brain activation patterns. Poldrack et al. (2013) adopted the Latent Dirichlet Allocation (LDA) method to identify topics of articles from the Neurosynth database and mapped these topics to brain activation data for discovering mechanisms of cognitive states. French et al. (2012) developed a co-occurrence-based method to extract brain regions and their relations from neuroscience articles. Alhazmi et al. (2018) extracted topic words based on word frequency and constructed relations between semantic spaces of topics and brain-activated regions by using correspondence analysis and hierarchical clustering. In order to decode cognitive states, these studies only extracted cognitive states and activated brain regions and their relations from neuroimaging articles. Main adopted text mining technologies are the frequency-based or probabilistic model-based topic learning methods and the clustering-based relation recognition methods. The obtained knowledge was a group of topic words and their relations. Because of adopting an open-domain task to extract unknown types of topics and relations, many general words were often included in topics, resulting in low-quality results of knowledge extraction. For example, neurosynth topic words include many general words, such as "using," "repeat," "asked," and domain general words, such as "magnetic resonance," "brain" (Abacha et al., 2017). Poldrack et al. (2013) had to use concepts in the Cognitive Atlas (Poldrack et al., 2011) to further filter neurosynth topic words. Such knowledge extraction oriented to the decoding of cognitive states cannot effectively characterize the whole research process for sharing "research objects."

In recent years, studies on neuroimaging text mining began to expend the knowledge extraction perspective from decoding cognitive states to characterize the whole research process. The adopted methods also change from topic modeling to named entity recognition. Ben Abacha et al. (2017) adopted the rule-based method and various conditional random field (CRF)-based methods to recognize fifteen functional neuroimaging entity categories, including gross brain anatomy, functional neuroanatomy, medical problem, stimuli, and responses, etc. Shardlow et al. (2018) recognized various entities, including brain regions, experimental values, neuron types, etc., by using active and deep learning, for curating researches in computational neuroscience. Riedel et al. (2019) completed a comprehensive evaluation on recognizing various entities related to the cognitive experiment, which are defined by the Cognitive Paradigm Ontology (Turner and Laird, 2012) and involved with behavioral domain, paradigm class, instruction, stimulus modality, etc., based on multiple corpus features and various classifiers, including Bernoulli naïve Bayes, k-nearest neighbors, logistic regression, and support vector classifier. Huangfu et al. (2020) extracted the term list of the neuroscientific research process from the PubMed Central database, for creating neuroscientific knowledge organization system (i.e., hierarchical terminology system). Zhu et al. (2020) proposed the hierarchical attentive decoding to extract a span of interests (i.e., terms about neuroscientific research process) from neuroscience articles, to predict research species. Whether it is the entity, method term, or research interest, the extracted knowledge in current studies of neuroimaging text mining is no longer limited to cognitive states and brain regions but extended to the whole research process, from the experiment to analysis. Furthermore, most of these studies transformed the open-domain topic modeling task to the close-domain entity recognition task with predefined entity/term/interest types from external domain knowledge, such as Cognitive Paradigm Ontology. By this kind of transformation of task definition, the shortcoming of general word noises in early studies of neuroimaging text mining can be effectively solved. Therefore, rapid growth neuroimaging articles, especially open access articles, and the maturing technologies of neuroimaging text mining provide a practical approach for continuous and rapid collection of neuroimaging provenance information.

However, for provenance-based research sharing, existing studies on neuroimaging text mining still have the following shortcoming. Words or phrases are the main form of extracted knowledge. Because of lacking the relation structure among words, these words or phrases can only show some fragmentary and key points of the research process and cannot provide a vivid process description. Hence, it is necessary to develop a new technology of neuroimaging text mining for research sharing-oriented neuroimaging provenance construction.

Biomedical Event Extraction

In the definition of ACE (Automatic Content Extraction) (Doddington et al., 2004), "event" is described as the occurrence of an action or the change of state. It includes an event trigger and multiple arguments with different roles. Event extraction is to obtain structured representations of events, so as to help answering the "5W1H" questions, including "who, when, where, what, why" and "how," of real-world events from numerous sources of texts (Xiang and Wang, 2019). Biomedical event extraction is to extract events from biomedical texts and has become one of the most actively researched areas in biomedical text mining (Chung et al., 2020). Since the first biomedical natural language processing shared task (BioNLP-ST) challenge in 2009, the fine-grained biomedical event extraction has received extensive attentions from academia and industry.

Different from news (Abera, 2020), finance (Zheng et al., 2019) and other fields, biomedical event extraction has two main challenges. Firstly, biomedical texts have the higher complexity. They often contain many abbreviations and long sentences with complex structures. Secondly, high-quality annotated corpora are lacking. Most of the existing biomedical event corpora are small scale and have serious class imbalance problems. In recent years, a large number of studies have tried to solve these two problems. Various embedding technologies, such as dependency

path embedding (Bjrne and Salakoski, 2018) and entity property embedding from external ontologies (Li et al., 2016), were developed to capture multi-aspect feature information for modeling complex biomedical texts. The hierarchical attention mechanism (Chen, 2019) was proposed to model the global document context for identifying biomedical event triggers from long biomedical sentences. Distance supervision (Araki and Mitamura, 2018), transfer learning (Chen, 2019), and other technologies have also been adopted to solve the problem of lack and imbalance of training samples.

In recent years, the knowledge base-oriented extraction task is becoming the research focus of biomedical event extraction. It evaluates the extraction systems or methods by measuring how much information content can be extracted from corpora (Deléger et al., 2016). Traditional pipeline models divided the event extraction task into multiple independent sub-tasks and resulted in various challenges, such as error propagation and lack of dependencies and interactions among subtasks. Therefore, the joint model becomes the inevitable choice of event extraction that oriented the knowledge base. Nguyen and Nguyen (2019) proposed a novel model to jointly perform predictions for entity mentions, event triggers, and arguments based on the shared hidden representations. Yu et al. (2019) developed an end-to-end model based on LSTM (Long Short-Term Memory) to optimize biomedical event extraction. Trieu et al. (2020) proposed an end-to-end neural nested event extraction model, named DeepEventMine, which can extract multiple overlapping directed acyclic graph structures from a raw sentence. Li et al. (2020) proposed a parallel multi-pooling convolutional neural network model to capture the compositional semantic features of sentences for improving biomedical event extraction on the MLEE (Multi-Level Event Extraction) dataset. Zhao et al. (2021) proposed a novel framework of reinforcement learning for the task of multiple biomedical event extraction, in which trigger identification and argument detection were treated as the main task and the subsidiary task.

Because of rich semantic information, the event is naturally more suitable to represent neuroimaging provenances, which characterize the research process of computational science, than isolated words. However, there are still two challenges in applying existing biomedical event extraction technologies, especially various joint models of biomedical event extraction, on research sharing-oriented neuroimaging provenance construction:

- Modeling the neuroimaging research process and result based on events. Existing biomedical event extraction tasks mainly focus on gene, protein, and disease-related events. As a new field of biomedical event extraction, neuroimaging event extraction lacks referential studies for task definition. Therefore, it is necessary to model the neuroimaging research process and results based on the perspective of events firstly. This is a challenging work because the definition of events needs to take into account both the demands of research sharing and the availability of knowledge in articles. The task type, content, and granularity of event extraction needs to be designed systematically by balancing various factors, including the task complexity, the capability of existing technologies, the importance and completeness of knowledge in research sharing, and the availability of knowledge in articles.
- Constructing the joint model of neuroimaging events in a few-shot learning scenario. Research sharing-oriented neuroimaging provenance construction is a typical task of knowledge base extraction and needs to adopt the joint model of event extraction. However, the complex network structure of joint model brings the higher requirements on the scale of training data. As a new field of event extraction, there are no existing annotated corpora for neuroimaging event extraction. Furthermore, neuroimaging event extraction in this study is involved with multi-domain entities for characterizing the research process of computational neuroscience. This also hinders the use of distance supervision (Araki and Mitamura, 2018) and transfer learning (Chen, 2019). Therefore, this is necessary to develop a new joint model of biomedical event extraction based on a small-scale "gold" corpus set annotated by domain experts.

CONTRIBUTIONS

The main contributions can be summarized as follows:

1. Firstly, this paper designs a provenance-guided approach for modeling the neuroimaging research process and result based on events. Considering the requirements of research sharing in open and FAIR neuroscience and availability of knowledge in neuroimaging articles, an improved brain informatics (BI) provenance model is defined based on NIDM. A mapping between the BI model and the definition of events is also proposed to obtain six categories of neuroimaging event-containing attributes for constructing research sharing-oriented neuroimaging provenances.
2. Secondly, this paper proposes a joint extraction model based on deep adversarial learning, called the adversarial training based neuroimaging event and attribute extraction model (AT-NeuroEAE), to extract the defined neuroimaging events. To the best of our knowledge, it is the first event extraction joint model containing attributes. Furthermore, the FreeAT (Free Adversarial Training)-based adversarial learning is introduced into the event joint extraction for improving the accuracy of event extraction in a few-shot learning scenario. This kind of optimization makes the proposed model more practical for computational neuroscience which lacks the "gold" labeled corpora.
3. Thirdly, a group of experiments were performed based on real data from the journal PLOS ONE. Experimental results show that, compared with existing event extraction models, the proposed model can more effectively and completely extract neuroimaging provenance information from neuroimaging articles based on a small-scale corpus set annotated by domain experts. This is a practical

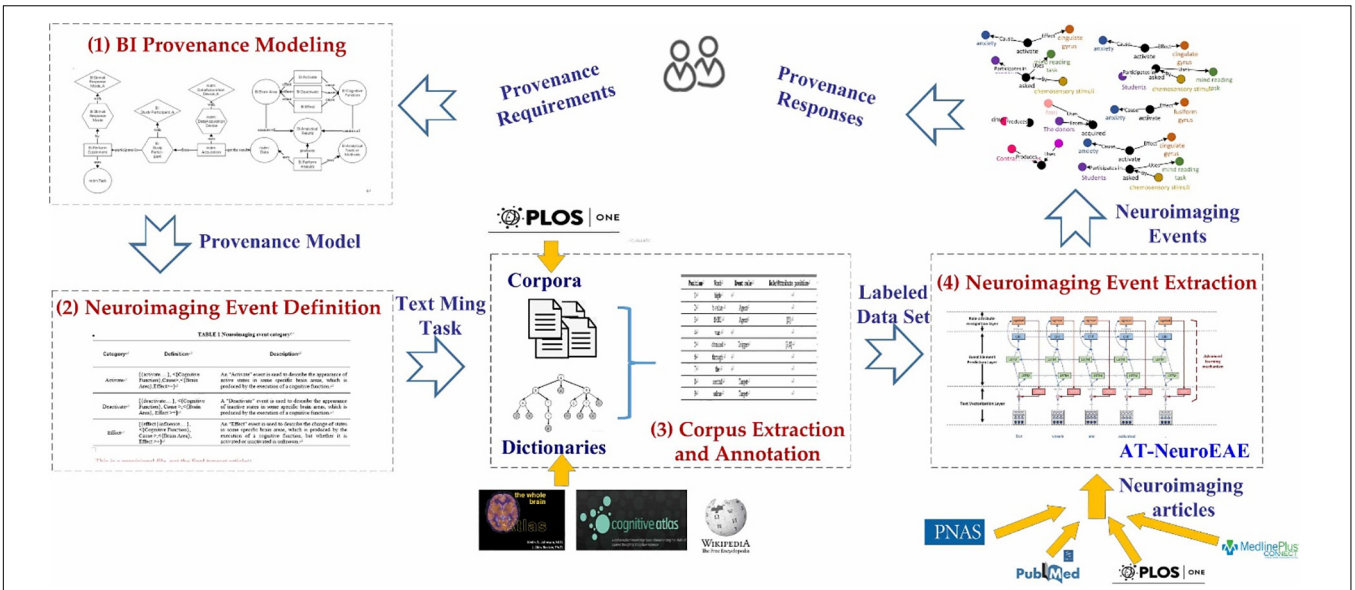


FIGURE 1 | The whole process of provenance information extraction. (1) Brain informatics provenance modeling: construct an improved BI provenance model to capture the provenance requirements of research sharing in open and FAIR neuroscience. (2) Neuroimaging event definition: define a group of neuroimaging events to transform the BI provenance model into text mining tasks. (3) Corpus extraction and annotation: construct a group of labeled corpora for model training and test. (4) Neuroimaging event extraction: develop the NeuroEAE model to extract defined neuroimaging events for meeting the provenance requirements in open and FAIR neuroscience.

and effective approach for large-scale and low-cost neuroimaging provenance construction.

MATERIALS AND METHODS

Overall Structure

This paper proposes a joint extraction model of events with attributes, called AT-NeuroEAE, for research sharing-oriented neuroimaging provenance construction. By using NeuroEAE, provenance information can be automatically and quickly extracted from published neuroimaging articles, especially open-access scientific articles. As shown in **Figure 1**, the whole process includes four steps. Details will be described in the subsequent subsections.

Brain Informatics Provenance Modeling

Computational neuroscience is a new application field of biomedical event extraction. In order to construct neuroimaging provenance by using biomedical event extraction, the first task needs to model the neuroimaging research process and result.

At present, NIDM is the most widely used neuroimaging provenance model and the representative work for open and FAIR Neuroscience (Abrams et al., 2021). However, as stated above, its element granularity is still incomplete for neuroimaging research sharing. It also contains a lot of model elements which aren't available in articles. In our previous studies (Chen et al., 2012; Chen and Ning, 2013), a BI provenance model has been proposed to describe the origin and subsequent processing of various human brain data in systematic BI studies. This study reconstructs the BI provenance model by considering both the importance of knowledge in research sharing and the availability of knowledge in articles.

Example 1: Participants were the same as in our previous study and comprised 18 unmedicated patients (age 39.8 years SD 12.8, 14 women). (Daniel, et al., 2014)

Example 2: Fourteen subjects with unilateral tinnitus were recruited at the University Medical Center Groningen, all without neurological and psychiatric history. (Lanting, et al., 2014)

FIGURE 2 | Diverse attributes of subjects in different articles. In example 1 (Daniel et al., 2014), the attributes of subjects include medication history, age, medical history, and gender. In example 2 (Lanting et al., 2014), the attributes of subjects include health condition and medical history.

Firstly, the newest version of BI provenance model (Sheng et al., 2019) is updated based on the widely used NIDM. All entities, activities and agents are replaced by NIDM classes, such as "nidm:AcquisitionObject," "nidm:Task," "nidm:Acquisition," and "nidm:DataAcquisitionDevice." This update ensures that the new BI provenance model follows the FAIR facets I1: "(meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation" and I2: "(meta) data use vocabularies that follow FAIR principles."

Secondly, according to the FAIR facet F2, "data are described with rich metadata" and the deficiency of NIDM in experiment description; the abstract classes of experimental design in NIDM is extended as the activity "BI:PerformExperiment" and three related elements, including the entity "nidm:Task" and the entity "BI:StimuliResponseMode." The entity "nidm:Task" and "nidm:StudyParticipant" are inherited from NIDM and used to indicate the task design type. The entity

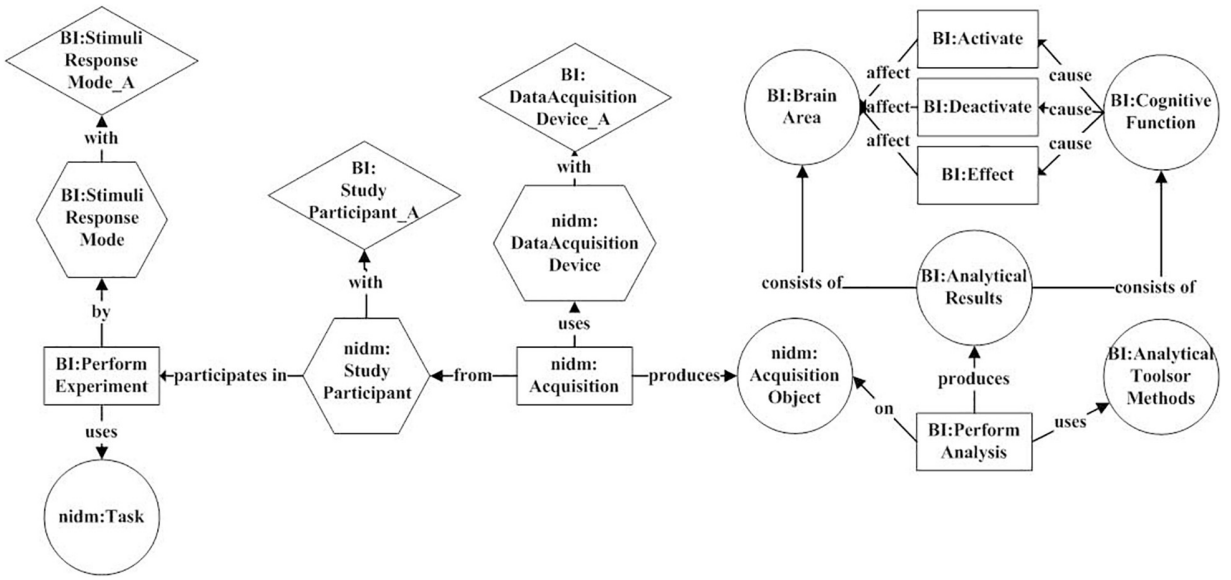


FIGURE 3 | The BI provenance model. It uses six activities (rectangle), including "BI:PerformExperiment," "nidm:Acquisition," "BI:PerformAnalysis," "BI:Activate," "BI:Deactivate," and "BI:Effect" to characterize experimental design, analytical process, and results. Six entities (circle), three agents (hexagon), and three attributes (diamond) are connected to these activities for describing various key factors in the research process.

"BI:StimuliResponseMode" are used to describe the experiment in more detail. The former describes the sensory stimuli or response of experiment and the latter describes the participant of experiment. They are two important factors for cognitive experiments and need to be considered in multiple key steps of computational neuroscience, including experimental design, data analysis, result interpretation, etc. Neuroimaging articles also often introduce them in the "Experiment" section. Furthermore, both the stimuli mode and the subject are involved with multi-aspect attribute information, such as the perception channel, the stimuli task category, age, gender, and occupation, but it is not sure which and how much attribute information will be described in an article. As shown in **Figure 2**, different articles may describe subjects by using different attributes, including medication history, age, gender, medical history, and other complex characteristics. Therefore, this study adds the general attribute entity to represent all attributes of the entity. For example, the entity "BI:StudyParticipant_A" is added into the model to represent all possible attribute information of subjects. Similarly, according to the FAIR facet F2 "data are described with rich metadata" and the deficiency of NIDM in result description; this study adds three categories of activities "BI:Activate," "BI:Deactivate," "BI:Effect" and two related elements, including the entities "BI:BrainArea" and "BI:CognitiveFunction," to describe the brain mechanism of cognitive states, which are core research findings in computational neuroscience.

Thirdly, considering both the importance and the availability of information, this study simplifies the description of NIDM about analytical tools or methods. NIDM contains many classes about parameters of analytical tools or methods. However, many parameters are generally not mentioned in the article. Therefore, this study uses an entity "BI:AnalyticalToolsorMethods," which is connected to the activity "BI:PerformAnalysis," to represent analytical tools or methods. Two entities "nidm:AcquisitionObject" and "BI:AnalyticalResults" are also connected to the activity "BI:PerformAnalysis" for representing the analytical process more clearly. Compared with the traditional class or concept tree, this kind of event-based description contains various roles within the event and can better describe the research process than the single subclass relation.

Figure 3 gives the reconstructed BI provenance model. It characterizes experimental design, analytical process, and results by four categories of elements: entity, activity, agent, and attribute.

Model-Guided Neuroimaging Event Definition

This study designs a model-guided approach to transform BI provenance model into text mining tasks. A group of neuroimaging events are defined by using the following mapping rules:

- Firstly, each activity and its directly connected elements (entities or agents) are divided into an independent activity unit.
- secondly, each activity unit is mapped to a meta event. The activity is mapped to the trigger words. The entities and agents are mapped to arguments. The relations between activity and elements are mapped to argument roles.

Based on the above rules, six categories of neuroimaging events can be obtained from the BI provenance model. **Tables 1**, **2** give the category definitions of events and corresponding arguments.

TABLE 1 | Neuroimaging event category.

Category	Definition	Description
Activate	[{Activate. . . }, <{cognitive function}, cause >, <{brain area}, affect > +]	An "Activate" event is used to describe the appearance of active states in some specific brain areas, which is produced by the execution of a cognitive function.
Deactivate	[{deactivate. . . }, < {Cognitive Function}, cause >, <{Brain Area}, affect > +]	A "Deactivate" event is used to describe the appearance of inactive states in some specific brain areas, which is produced by the execution of a cognitive function.
Effect	[{effect | influence. . . }, < {Cognitive Function}, cause > , < {Brain Area}, affect > +]	An "Effect" event is used to describe the change of states in some specific brain areas, which is produced by the execution of a cognitive function, but whether it is activated or inactivated is unknown.
Perform experiment	[{perform | complete. . . }, < {Study Participant (Study Participant_A)*}, participates in > *, < {Task}, uses > +, < {Stimuli Response Mode (Stimuli Response Mode_A)*}*, by >]	A "Perform Experiment" event is used to describe a neuroimaging research action by a group of study participants, i.e., subjects do one or several experimental tasks by some kind of stimuli response mode. Study participants often have some attributes, such as age, gender, and medical history. The stimuli response mode is also involved with some features, such as the perception channel, the stimuli task category.
Acquisition	[{assess | acquire | obtain...}, < {Acquisition Object}, produces > *, < {Data Acquisition Device (Data Acquisition Device_A)*}, uses > +, < {Study Participant (Study Participant_A)*}, from > *]	The "Acquisition" event is used to describe a research action that a data acquisition device with relevant parameters produces physiological and psychological data from study participants.
Perform analysis	[{perform | complete. . . }, < {Analytical Results}, produces > +, < {Acquisition Object}, on > *, < {Analytical Tools or Methods}, uses > +]	The "Perform Analysis" event is used to describe a research action that the analytical tool or method is used on physiological and psychological data to produce a group of analytical results, i.e., brain responses, such as Default Mode Network (dmn).

"" indicates that the element may occur zero or more times and "+" indicates that the element may occur one or more times.*

TABLE 2 | Argument categories.

Argument category	Description
Cognitive function (COG)	The cognitive function is an ability of the human brain to process information and used to denote the brain function implied by brain responses in computational neuroscience research.
Brain area (BRI)	The brain area is an anatomical region of the cerebral cortex and used to mark the occurrence location of brain response in computational neuroscience research.
Data acquisition device (ACQ)	The data acquisition device is a kind of brain testing equipment used in computational neuroscience research.
Stimuli response mode (SEN)	The stimuli response mode is used to denote the sensory channel of stimulus presentation in computational neuroscience research.
Study participant (STP)	The study participant is a participant in computational neuroscience research and recorded for behavioral or brain physiological data.
Task (TSK)	The experimental task is a task (e.g., questions, games, etc.) which is performed by the study participant in computational neuroscience research.
Acquisition object (AOB)	The acquisition object is a kind of physiological and psychological data which are collected by the data acquisition device in computational neuroscience research.
Analytical tools or methods (TOL)	The analytical tool or methods is a mining algorithm or tool which is used to analyze experimental data in computational neuroscience research.
Analytical results (RLT)	The analytical results are a series of brain responses which are mined from experimental data in computational neuroscience research.

In this study, the definition of events is represented as trigger + parameter structure (Sun et al., 2017). For example, an "Acquisition" event can be represented as follows:

Event Acquisition

= [trigger, < argument1(attribute)#, role1 > #, < argument2(attribute)#, role2 > #...]

= [{assess | acquire | obtain...}, < {Acquisition Object}, produces > *, < {Data Acquisition Device (Data Acquisition Device_A)*}, uses > +, < {Study Participant (Study Participant _A)*}*, from >]

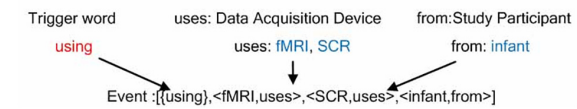


FIGURE 4 | An example of an "Acquisition" event from the article (Mutschler et al., 2016). This event consists of a trigger word "using," two "Data Acquisition Device" category of arguments "fMRI" and "SCR," and one "Study Participant" category of argument "infant."

The above expression shows that an "Acquisition" event consists of a trigger word and three categories of arguments. "*" indicates that the element may occur zero or more times and "+" indicates that the element may occur one or more times. The trigger word may be one of "assess," "acquire," "obtain," etc. Argument 1 belongs to the "Acquisition Object" category and is

TABLE 3 | The number of terms in dictionaries.

Element type	Element	Term number	Element	Term number
Trigger word	Acquisition	65	Effect	145
	Perform experiment	62	Deactivate	20
	Perform analysis	227	Activate	12
Argument	BRI	576	SEN	52
	COG	812	STP	97
	AOB	48	TOL	171
	ACQ	57	TSK	814
	RLT	28		
Argument	Data acquisition device_A	33	Study participant_A	7
	Stimuli response mode_A	24		

used to describe the results produced by the "Acquisition" event. Argument 2 belongs to the "Data Acquisition Device" category and is used to describe the device used in the "Acquisition" event. Argument 3 belongs to the "Study Participant" category and is used to describe the object of the "Acquisition" event. Both Argument 2 and Argument 3 have zero or several attributes. **Figure 4** gives an example of an "Acquisition" event in a neuroimaging article.

Corpus Extraction and Annotation

Corpora were crawled from the journal PLOS ONE. We searched the articles based on the keywords "fMRI" or "functional magnetic resonance imaging" or "functional MRI" in their abstracts, and set the published time from July 18, 2014, to July 18, 2019. A total of 677 articles were obtained. Based on these articles, event mentions were extracted by using rule matching. The process can be described as follows.

The first step is dictionary construction. Seven categories of argument dictionaries, including "Brain Area," "Cognitive Function," "Study Participant," "Stimuli Response Mode," "Task," "Analytical Tools or Method," and "Analytical Results" came from our previous study (Sheng et al., 2020). We collected candidate trigger words and arguments based on the co-occurrence with these seven categories of arguments, and then manually filtered them to construct the trigger word dictionaries and other categories of argument dictionaries. Similarly, three categories of argument attribute dictionaries were constructed based on the co-occurrence and manual filtering. **Table 3** gives the number of terms in each dictionary.

The second step is mention extraction. Based on the dictionaries and event definitions, event mentions were extracted by using the following rule:

> *Rule1: If an event trigger word and two arguments appear in the same sentence, this sentence can be marked as a complete event.*

However, Rule 1 cannot extract enough the "Deactivate" event. We defined another more relaxed rule:

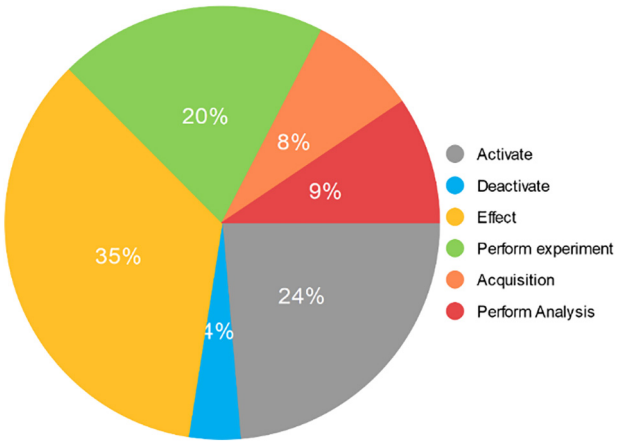


FIGURE 5 | The distribution of event mentions in the experimental data set. It consists of 3331 event mentions extracted from 677 neuroimaging articles. The "Activate" category includes 788 mentions and accounts for 24% of the total. The "Deactivate" category includes 128 mentions and accounts for 4% of the total. The "Effect" category includes 1169 mentions and accounts for 35% of the total. The "Perform experiment" category includes 665 mentions and accounts for 20% of the total. The "Acquisition" category includes 266 mentions and accounts for 8% of the total. The "Perform Analysis" category includes 315 mentions and accounts for 9% of the total.

> *Rule2: If a "Deactivate" event trigger word and a "BRI" argument appear in the same sentence, this sentence can be marked as a "Deactivate" event.*

Based on Rules 1 and 2, a neuroimaging event mention set can be obtained as the experimental data set. It includes 788 "Activate" event mentions, 128 "Deactivate" event mentions, 1169 "Effect" event mentions, 665 "Perform Experiment" event mentions, 266 "Acquisition" event mentions, and 315 "Perform Analysis" event mentions. The distribution of event mentions in the experimental data set is shown in **Figure 5**.

In the training dataset, the annotation of triggers and arguments adopts the "B/I/O-category abbreviation" mode. The "B/I/O" is the BIO annotation mode (Soomro et al., 2017). "B" indicates the beginning of entity. "I" stands for the middle or end, and "O" stands for other, which is used to mark irrelevant characters. Category abbreviations are involved with argument categories and trigger categories. The entity annotation of attributes adopts the "B/I/O-category abbreviation_A" model. **Table 4** gives an annotation example. The labels "BRI" and "AOB" indicate the argument categories "Brain Area" and "Acquisition Object," respectively. The label "Acq" indicates the trigger category "Acquisition." The label "O" on the word "high" indicates "high" does not belong to any entity. The label "B-AOB" on the word "fMRI" indicates "fMRI" is the beginning of the "Acquisition Object" argument.

The annotation of event roles and attributes is based on trigger words. As shown in **Table 5**, the first column is the position and the second column is the corresponding word. If the word is a trigger word, its third column is marked "Trigger" and its fourth column is marked with the form of "[$ArgPos_1$, $ArgPos_2$,. . .,

TABLE 4 | Event element annotation.

				Annotation example				
High	***b*-value**	**fMRI**	**Was**	**Obtained**	**Through**	**The**	**Central**	**Sulcus**
O	O	B-AOB	O	B-Acq	O	O	B-BRI	I-BRI

TABLE 5 | Event role and attribute annotation.

Position	Word	Event role	Role/attribute position
1	High		
2	*b*-value	Produces	
3	fMRI	Produces	[2]
4	Was		
5	Obtained	Trigger	[3,8]
6	Through		
7	The		
8	Central	From	
9	Sulcus	From	

$ArgPos_n$,. . .]" in which "$ArgPos_n$" is the starting position of the corresponding argument in $role_n$. If the word is an argument, its third column is marked by the corresponding role category label. And if this argument has attributes, its fourth column is marked with the form of "[$AttPos_1$, $AttPos_2$,. . ., $AttPos_n$,. . .]" in which "$AttPos_n$" is the starting position of its $attribute_n$.

Neuroimaging Event Extraction Based on AT-NeuroEAE

In order to realize the event extraction containing attributes, this study proposes a joint event extraction model based on deep adversarial learning, called AT-NeuroEAE. The multi-layer joint extraction model of neuroimaging events with attributes is constructed to synchronously realize the prediction of event elements, including trigger, arguments, and attributes and the extraction of element relations, including argument roles and argument attributes, by the end-to-end mode. The adversarial learning based on FreeAT is combined with this joint model to realize the event extraction in a few-shot learning scenario. **Figure 6** gives the whole structure of the model, which consists of the text vectorization layer, the event element prediction layer, the role-attribute recognition layer, and the adversarial learning mechanism.

In this experiment, we set the dimension of the word vector to 25, the number of iterations to 200, the dimension of the LSTM hidden layer to 64, the number of LSTM layers to 3, the dropout to 0.9, the learning rate to 1e-3, the activation function to tanh, and the adversarial disturbance parameter alpha to 0.01. Parameter optimization used the Adam algorithms.

Text Vectorization Layer

The text vectorization layer encodes sentences as textual vectors for the input of model. Suppose that there is a sentence $s = [w_1, w_2, ...w_n]$. For each word w_i, three categories of feature vectors are constructed based on lexical units, case features, and domain terminology dictionaries, respectively.

- Word vector. The word vector is learned to contain as much semantic and grammatical information as possible (Pennington et al., 2014). This study adopts the Glove word vector model which was trained on 6 billion words of Wikipedia and web texts (Pennington et al., 2014).
- Case feature vector. Neuroimaging event mentions are involved with a large number of domain terms with capitalized abbreviated names, such as fMRI. In order to capture this kind of corpus features, this study constructs a one-hot case feature vector including six dimensions "numeric," "allLower," "allUpper," "initialUpper," "mainly_numeric," "contains_digit," and "other."
- Terminology dictionary vector. Neuroimaging event elements involve nine types of important domain entities shown in **Table 2**. This study constructs the one-dimensional terminology dictionary vector (Xu et al., 2019) to capture this kind of domain corpus features. Nine term dictionaries are collected for these nine categories of domain entities firstly, and then a label list is constructed base on the "B/I/O-entity category abbreviation." When w_i matches any term in the dictionaries, a label index is set as the dimensional value of terminology dictionary vector of w_i.

Concatenating these three categories of feature vectors (Zhang et al., 2019), the combined word vector of w_i can be obtained as follows:

$$v_i = [v_w, v_c, v_t] \quad (1)$$

where v_w, v_c, and v_t are the corresponding word vector, case feature vector, and terminology dictionary vector of w_i.

Event Element Prediction Layer

The event element prediction layer predicts the potential event elements (i.e., argument, attribute, or trigger word) by using the BiLSTM-CRF model. BiLSTM is used to model context information of sentences based on input text vectors. The process can be defined as follows:

$$h_i = \left[\overrightarrow{LSTM}(v_i), \overleftarrow{LSTM}(v_i)\right], \ i \in [1,n] \quad (2)$$

where, v_i is the combination vector of w_i, $\overrightarrow{LSTM}$ is the forward hidden layer output of LSTM, $\overleftarrow{LSTM}$ is the backward hidden layer output of LSTM, and h_i is a word representation of each word token.

CRF is used to predict the element category and boundary. Event element prediction is a named entity recognition task which can be formulated as a sequence labeling problem. Based

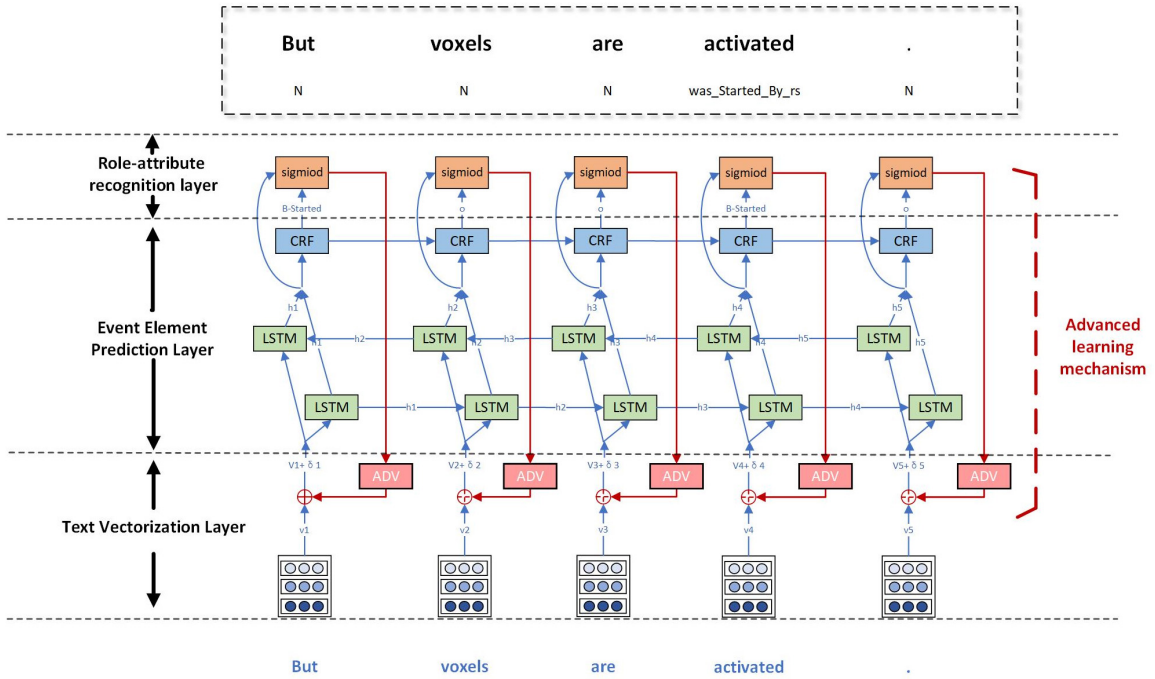


FIGURE 6 | The AT-NeuroEAE model. The text vectorization layer encodes sentences as textual vectors based on lexical units, case features, and domain terminology dictionaries. The event element prediction layer predicts the potential event elements by using the BiLSTM-CRF model. The role-attribute recognition layer identifies the role and attribute of argument by using the sigmoid function. The adversarial learning mechanism adds small and persistent disturbances to the input of joint model for improving the robustness and generalization of the model. BiLSTM: bi-directional long short-term memory; CRF: conditional random fields; adv: adversarial learning.

on the BIO annotation mode, each event element often consists of multiple sequential tokens. We calculate the score of each token w_i for each event element tag:

$$S^{(e)}\left(h_i\right) = V^{(e)} f\left(U^{(e)} h_i\right), \ \mathrm{i} \in [1,n] \tag{3}$$

where the superscript (e) means event element prediction; $f(\cdot)$ is the activation function; $V^{(\mathrm{e})} \in \mathbb{R}^{\mathrm{p}=l}$, $U^{(\mathrm{e})} \in \mathbb{R}^{2d=l}$, and l are the width of LSTM layer; and d is the hidden size of LSTM layer. The linear-chain CRF score of the s is defined as:

$$S\left(y_1{}^{(e)}, y_2{}^{(e)}, \ldots, y_n{}^{(e)}\right) = \sum_{i=0}^{n} s_{i,y_i{}^{(e)}}^{(e)} + \sum_{i=1}^{n-1} T_{y_i{}^{(e)}, y_{i+1}{}^{(e)}},$$

$$\mathrm{i} \in [1, \mathrm{n}] \tag{4}$$

where $y_i \in \mathrm{Y}$ is the tag of w_i, s_{i,y_i} is the predicted score of w_i given a tag y_i, and T is the transition matrix.

Then the probability of a given tag sequence on all possible tag sequences of the input sentence *sen* is defined as:

$$Pr = \left(y_1{}^{(e)}, y_2{}^{(e)}, \ldots, y_n{}^{(e)} \,\middle|\, sen\right) \frac{e^{S\left(y_1{}^{(e)}, y_2{}^{(e)}, \ldots, y_n{}^{(e)}\right)}}{\sum_{\widetilde{y_1}{}^{(e)}, \ldots, \widetilde{y_n}{}^{(e)}} e^{s\left(\widetilde{y_1}{}^{(e)}, \ldots, \widetilde{y_n}{}^{(e)}\right)}} \tag{5}$$

In the process of training, this study minimizes the cross-entry loss $\mathcal{L}_{NER}$:

$$\mathcal{L}_{NER} = \sum_{i=0}^{n} -logPr(y_i|w_i;\theta) \tag{6}$$

Role-Attribute Recognition

The role-attribute recognition layer identifies the role and attribute of argument by using the sigmoid function. Recognizing the attribute of argument is the one-to-one relation extraction between the argument and its attribute, but recognizing the role of argument is the one-to-n relations because the trigger word is usually corresponding to multiple arguments in an event. Hence, this paper models the role recognition task as a multi-label head selection problem (Zheng et al., 2017), i.e., the relation extraction between one trigger word and two or more arguments.

As shown in **Figure 4**, the role-attribute recognition layer receives the hidden state of LSTM from the event element prediction layer. The predication score between tokens w_i and w_j given a role-attribute tag r_k can be calculated as follows:

$$s^{(r)}\left(h_i, h_j, r_k\right) = V^{(r)} f\left(U^{(r)} h_i + W^{(r)} h_j\right) \tag{7}$$

where the superscript (r) means role-attribute recognition, $f(\cdot)$ is the activation function, h_i is the output of the hidden state of LSTM corresponding to w_i, $V^{(r)} \in \mathbb{R}^l$ l is the width of the LSTM layer, $U^{(r)} \in \mathbb{R}^{l*2d}$, $W^{(r)} \in \mathbb{R}^{l*2d}$, and d is the hidden size of LSTM layer. The probability of token w_j to be selected as the head of token w_i with the role-attribute tag r_k between them can be calculated as follows:

$$Pr\left(head = w_j,\ tag = r_k \,\middle|\, w_i, \theta\right) = \sigma(s(m_j, m_i, r_k)) \tag{8}$$

where θ is a set of parameters and σ(·) is the sigmoid function.

In the process of training, this paper minimizes the cross-entry loss $\mathcal{L}_{RAR}$:

$$\mathcal{L}_{RAR} = \sum_{i=0}^{n} \sum_{j=0}^{m} -Pr\left(head = y_{i,j}, relation = r_k | w_i\right) \quad (9)$$

where y_i is the ground truth vector and r_k is the role-attribute tag, and $y_i \subseteq sen$ and $r_k \subseteq \mathrm{R}$ and m ($m < n$) are the number of associated heads (i.e., roles or attributes) of w_i. During decoding, the most probable heads and roles/attributes are selected using threshold-based prediction.

Finally, for the joint event extraction task, the final objective is calculated as $\mathcal{L}_{NER}+\mathcal{L}_{RAR}$. This study minimizes this objective function in the process of training of joint model.

Adversarial Learning Mechanism

The adversarial learning mechanism adds small and persistent disturbances to the input of the joint model. After the adversarial training with disturbances, the output distribution is consistent with the original distribution, but the robustness and generalization of the model can be improved (Wang et al., 2018).

This study adopts the perturbation strategy of FreeAT proposed by Shafahi et al. (2019). As shown in **Figure 5**, repeat K times for each sample continuously to find the optimal perturbation. The gradient of the previous step is multiplexed when calculating the perturbation. Finally, the overall iteration is divided by K to greatly reduce the speed problem caused by the inner iteration. The approximation of the perturbation is defined as:

$$r_{adv\ t+1} = r_{adv\ t} + \in \frac{\nabla_w \mathcal{L}(w; \widehat{\theta})}{||\nabla_w \mathcal{L}(w; \widehat{\theta})||} \quad (10)$$

where, r_{adv} is the perturbation, $\widehat{\theta}$ is a copy of the current model parameter, $t \subseteq \mathrm{K}$, $\in = \alpha\sqrt{D}$ is a hyperparameter, α is the factor and D is the dimension of the embeddings v_i, and t is the number of inner iterations.

Finally combining the original examples and the confrontation examples for model adversarial training, the final loss of joint model is as follows:

$$\mathcal{L}\left(w, \widehat{\theta}\right) = \mathcal{L}_{joint}(w; \widehat{\theta}) + \mathcal{L}_{joint}(w + r_{adv}; \widehat{\theta}) \quad (11)$$

where $\widehat{\theta}$ is the current value of the model parameter.

Evaluation

The precision rate P, the recall rate R, and the F_1 value (Curiskis et al., 2019) are adopted to evaluate experimental results. They can be calculated as follows:

$$P = \frac{|TP|}{|TP| + |FP|} \quad (12)$$

$$R = \frac{|TP|}{|TP| + |FN|} \quad (13)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

In this study, P, R, and F_1 values are used to evaluate the results of subtasks, but only F_1 values are used to evaluate the overall results of event extraction.

RESULTS

Baseline Methods

The CNN-BiLSTM-PCNN (Convolutional Neural Network-Bi-directional Long Short-Term Memory-Pulse Coupled Neural Network) pipeline model (Zheng et al., 2017) and the GCN (Graph convolution Network) joint model[9] are used in the two control experiments. The CNN-BiLSTM-CNN pipeline model uses the CNN-BiLSTM model for the subtask of event element prediction, and the PCNN model for the subtask of role attribute recognition.

In many competitions, these models have been proved to be stable and relatively optimal for various biomedical named entity recognition tasks and relation extraction tasks. The GCN joint model was used on multiple shared tasks of biomedical event extraction, including BB3 (Bacteria Biotope) event data set (Lever and Jones, 2016), and SeeDev (Plant Seed Development) event data set (Agirre et al., 2019), and has obtained good experimental results.

In this study, these two models used their original parameters firstly, and then we adjusted parameters referring to some classic articles (Kip and Welling, 2016; Ma and Hovy, 2016; Zhang et al., 2019). Based on the results, a group of optimal parameters were chosen as follows:

- The CNN-BILSTM-PCNN pipeline model: For the event element prediction subtask, the dimension size of the word vector was set at 100. The epoch number was set at 50, the convolution width at 3, the CNN output size at 30, the dimensional number of LSTM hidden layer at 200, the mini-batch size at 9, and the dropout at 0.5. The Adam algorithm was used to optimize parameters. For the role-attribute recognition subtask, the dimension size of the word vector was set at 50, the convolution width at 3, the dimensions of the two hidden layers at 200 and 100, respectively, and the learning rate at 0.01.
- The GCN joint model: The dimension size of the word vector was set at 300, the dimension size of the Hidden layer at 200, the epoch number at 50, dropout of GCN at 0.5, dropout of word at 0.5, and the batch size at 16.

Results Analysis

This study adapted fivefold cross validation to improve the objectivity of results. Each category of event mentions was randomly divided into five equal parts. In order to simulate the few-shot learning scenario, each experiment used one part as the training data set and the remaining four parts as the test data set. After five experiments, the average results were taken as the final results.

[9]https://github.com/wangmm88/BioNLP-ST-2016_BB3-event

TABLE 6 | F_1 values of event extraction.

Event category	CNN-BiLSTM-PCNN	GCN	NeuroEAE	AT-NeuroEAE
Acquisition	0.588	0.542	0.689	0.734
Perform analysis	0.448	0.596	0.547	0.671
Perform experiment	0.368	0.429	0.612	0.853
Effect	0.569	0.576	0.690	0.692
Activate	0.614	0.682	0.809	0.756
Deactivate	0.922	0.952	0.974	0.608

Table 6 gives F_1 values of all models. It can be seen that the CNN-BiLSTM-PCNN pipeline event model lags behind the other three models in five categories of events except "Acquisition." This shows the necessity of joint model in the complex text mining task, such as event extraction. The proposed NeuroEAE model is superior to the GCN joint model in all event categories except the "Analytical results" event. This proves the validity of proposed model structure for extracting neuroimaging event containing attributes.

The proposed AT-NeuroEAE model introduces adversarial learning into NeuroEAE. It achieved the better results on most of event categories than other models. This shows the validity of adversarial learning in the few-shot learning scenario. However, AT-NeuroEAE loses to NeuroEAE on the "Activate" and "Deactivate" events. This may be due to the fact that its structure and elements are relatively single compared with other event categories. As the conclusion of computational neuroscience research, the "Activate" and "Deactivate" events are often described by using the simple and clear languages in articles. For example, "We previously showed that a common network of brain regions was activated for past and future thinking in healthy older adults" (Rodolphe et al., 2014) and "This cortical area was found to be activated in studies where participants had to perform a task of face vs. non-face recognition" (Prehn-Kristensen et al., 2009). Under the same degree of adversarial disturbance of the model, such a simple event structure is easy to produce the collapse problem, which makes the model unable to further learn. This shows that the increasing disturbance can lead to the reduction of model accuracy under the simple event extraction. For other event categories, whether it is the "Analytical Results" event with the complex structure or the "Effect" event with more trigger word instances, the model has advantages by using adversarial learning. Hence, the proposed AT-NeuroEAE is an effective model to extract neuroimaging events from complex event mentions, which are common in neuroimaging articles. If the event is simpler, the NeuroEAE model is the better choice. Extracting neuroimaging event-containing attributes can be realized by flexibly using these two kinds of models.

Table 7 gives the F_1 values of each experiment in fivefold cross validation. As shown in this table, the results are stable in each experiment. Our experiments were performed on the personal computer with CPU Intel Core 9th Generation i7 and Graphics Card NVIDIA GeForce GTX 1650. The epoch number was set at 100. **Table 8** gives the running time of each experiment. The average running time of model training is almost 4.8 h and the average running time of model test is almost 6.7 s. This shows that, although the model training takes a long time, the speed of event extraction is still relatively fast and can meet the requirements of rapid extraction of provenances.

DISCUSSION

Results Analysis on Subtasks

Neuroimaging event extraction consists of two subtasks: event element prediction and role-attribute recognition. This section will analyze results of each subtask separately to understand and evaluate the proposed AT-NeuroEAE model in depth.

Table 9 gives the results of event element prediction, which are involved with three kinds of elements, i.e., trigger words, arguments, and attributes. The full names of arguments are mentioned in **Table 2**. The "abbreviation_A," such as DAT_A, denotes the attribute of the corresponding argument. In this study, the prediction task of attribute is most ambiguous because the general attribute indicates all of possible actual attributes, as stated above. The prediction task of arguments is clearly defined but each argument category is often involved with a large number of entity instances. The prediction task of trigger words is clearest and simplest. Each category of trigger words only involved limited words. Comparing prediction results on these three kinds of elements, we can see that the clearer and simpler the task, the better the result. Generally, all elements are recognized well. F_1 values of the AT-NeuroEAE model are almost above 0.9. Even the worst model CNN-BiLSTM-PCNN, its F_1 values on all categories of elements are more than 0.7. This shows that the existing deep learning models have been able to effectively solve the task of named entity recognition, as stated in Armelle et al. (2014).

Comparing AT-NeuroEAE and NeuroEAE, most F_1 values are improved and proved the effectiveness of adversarial learning. The negative effects of adversarial learning mainly appear in attribute prediction. This shows that the ambiguous task definition may not be suitable for adversarial learning.

Table 10 gives the results of role-attribute recognition, which are involved with two aspects of relations, i.e., element roles and element-attribute relations. Obviously, adversarial learning improves the F_1 values of role-attribute recognition for the majority of roles and attributes. The F_1 value of "Analysis-Agent" role is significantly improved after adding adversarial learning. The main reason is that the corresponding "Perform Analysis" event has the complex structure. This confirms the above judgment that the adversarial learning is more suitable for complex tasks.

In addition, we can see that the F_1 values of the "Activate-affect" role and "Activate-cause" role in the "Activate" event decrease after adding adversarial learning. As stated above, the "Activate" event often has the simple and clear language structure because it is the research conclusion. **Table 9** shows that the F_1 values of the element "Brain Area" (BRI) in this event

TABLE 7 | F_1 values in fivefold cross validation.

Event category	F_1-values					
	1	2	3	4	5	Average
Acquisition	0.700	0.720	0.768	0.763	0.722	0.734
Perform analysis	0.676	0.674	0.699	0.610	0.696	0.671
Perform experiment	0.889	0.837	0.863	0.871	0.808	0.853
Effect	0.724	0.686	0.696	0.734	0.618	0.692
Activate	0.772	0.756	0.740	0.781	0.728	0.756
Deactivate	0.639	0.613	0.608	0.600	0.583	0.608

TABLE 8 | Running times in fivefold cross validation.

Experiment	1	2	3	4	5	Average
Model training (s)	17194.8	17601.7	17331.0	17425.7	17191.0	17348.8
Model test (s)	7.46	5.95	6.79	6.98	6.18	6.7

TABLE 9 | Event element prediction.

Element		CNN-BiLSTM-PCNN			NeuroEAE			AT- NeuroEAE		
		P	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1
Trigger word	Acquisition	0.864	0.888	0.876	0.951	0.925	0.938	0.972	0.972	0.972
	Perform experiment	0.841	0.803	0.821	0.851	0.802	0.826	0.915	0.915	0.915
	Analytical results	0.879	0.79	0.832	0.981	0.879	0.927	0.843	0.931	0.885
	Effect	0.958	0.846	0.898	0.956	0.945	0.951	0.951	0.951	0.951
	Deactivate	0.821	1	0.902	1	0.958	0.978	1	1	1
	Activate	1	0.985	0.993	0.889	1	0.941	0.971	0.985	0.978
Argument	BRI	0.863	0.871	0.867	0.929	0.908	0.919	0.921	0.908	0.915
	COG	0.744	0.703	0.723	0.769	0.704	0.735	0.861	0.789	0.824
	AOB	0.941	0.899	0.92	0.975	0.91	0.941	0.988	0.943	0.965
	ACQ	0.962	0.962	0.962	0.897	1	0.897	1	1	1
	RLT	0.963	0.897	0.929	1	0.896	0.945	0.964	0.931	0.947
	SEN	0.972	0.648	0.778	0.895	0.781	0.834	0.959	0.854	0.903
	STP	0.744	0.727	0.736	0.851	0.909	0.879	0.811	0.977	0.886
	TOL	0.776	0.844	0.809	0.864	0.711	0.78	0.907	0.867	0.886
	TSK	0.982	0.873	0.924	0.936	0.936	0.937	0.953	0.968	0.961
Argument attribute	ACQ_A	0.667	0.905	0.776	0.944	0.809	0.871	0.9	0.857	0.878
	SEN_A	0.744	0.806	0.773	0.794	0.75	0.771	0.965	0.778	0.862
	STP_A	0.962	0.833	0.893	0.964	0.9	0.931	1	0.966	0.983

also decrease after adding adversarial learning. This confirms the above judgment again that the adversarial learning is not suitable for simple tasks. The F_1 values of another event element "Cognitive Function" (COG) don't decrease mainly because "Cognitive Function" is most complex element category and has thousands of entity instances (Poldrack et al., 2011) in different forms. The F_1 values of two categories of attribute relations "ACQ-Attribute" and "SEN-Attribute" also decrease after adding adversarial learning. This confirms the above judgment again that adversarial learning is not suitable for ambiguous tasks.

Comparing **Tables 9**, **10**, we can see that in the first subtask, the results of CNN-BiLSTM-PCNN pipeline model are similar to our two joint models, but there is a big gap in the second subtask. The reason is the existence of cascading errors. As shown in **Table 11**, when cascading errors are removed manually, CNN-BiLSTM-PCNN can also achieve good results. This proves the validity of joint models.

Result Comparison With Existing Studies on Neuroimaging Text Mining

As stated in section "Neuroimaging Text Mining," existing studies on neuroimaging text mining can be divided into neuroimaging topic modeling and neuroimaging named entity recognition. This section will compare the proposed AT-NeuroEAE with these two kinds of neuroimaging text mining, respectively.

LDA is the most commonly used method for neuroimaging topic modeling and has been applied to many previous studies (French et al., 2012; Alhazmi et al., 2018). Different from

TABLE 10 | Role-attribute recognition.

Event role	CNN-BiLSTM-PCNN			NeuroEAE			AT- NeuroEAE		
	P	*R*	F_1	*P*	*R*	F_1	*P*	*R*	F_1
Acquisition-uses	0.597	0.74	0.661	0.708	0.667	0.687	0.766	0.706	0.735
Acquisition-produces	0.558	0.898	0.688	0.772	0.809	0.791	0.8	0.762	0.781
Perform analysis-on	0.571	0.667	0.615	0.681	0.687	0.684	0.729	0.794	0.761
Perform analysis-produces	0.5	0.692	0.581	0.686	0.528	0.597	0.687	0.632	0.658
Perform analysis-uses	0.7	0.333	0.451	0.667	0.5	0.571	0.642	0.843	0.729
Perform experiment-by	0.471	0.696	0.562	0.56	0.583	0.571	0.619	0.541	0.578
Perform experiment-participates in	0.3	0.462	0.364	0.416	0.357	0.384	0.462	0.428	0.444
Perform experiment-uses	0.511	0.649	0.571	0.56	0.736	0.636	0.591	0.684	0.634
Effect-affect	0.75	0.788	0.768	0.711	0.821	0.762	0.832	0.806	0.818
Effect-cause	0.649	0.649	0.649	0.677	0.677	0.677	0.731	0.686	0.708
Deactivate-affect	0.875	0.966	0.918	1	1	1	1	1	1
Deactivate-cause	0.905	0.95	0.918	0.952	0.952	0.952	1	1	1
Activate-affect	0.862	0.812	0.836	0.833	0.864	0.848	0.861	0.765	0.81
Activate-cause	0.566	0.625	0.594	0.676	0.588	0.625	0.667	0.549	0.602
ACQ-attribute	0.594	0.92	0.867	0.956	0.709	0.814	0.806	0.806	0.806
SEN-attribute	0.385	0.385	0.385	1	0.526	0.689	0.9	0.346	0.5
STP-attribute	1	0.815	0.898	0.692	0.62	0.654	0.772	0.586	0.667

TABLE 11 | Comparison of pipeline models with and without cascading errors.

Event role	CNN-BiLSTM-PCNN (no cascading errors)			CNN-BiLSTM-PCNN		
	P	*R*	F_1	*P*	*R*	F_1
Acquisition-uses	0.657	0.98	0.951	0.597	0.74	0.661
Acquisition-produces	0.612	1	0.759	0.558	0.898	0.688
Perform analysis-on	0.695	0.842	0.761	0.571	0.667	0.615
Perform analysis-produces	0.692	0.857	0.765	0.5	0.692	0.581
Perform analysis-uses	1	0.478	0.647	0.7	0.333	0.451
Perform experiment-by	0.469	0.958	0.63	0.471	0.696	0.562
Perform experiment-participates in	0.423	0.785	0.55	0.3	0.462	0.364
Perform experiment-uses	0.692	0.947	0.799	0.511	0.649	0.571
Effect-affect	0.886	0.921	0.903	0.75	0.788	0.768
Effect-cause	0.783	0.775	0.779	0.649	0.649	0.649
Deactivate-affect	0.906	1	0.951	0.875	0.966	0.918
Deactivate-cause	0.869	1	0.93	0.905	0.95	0.918
Activate-affect	0.975	0.962	0.968	0.862	0.812	0.836
Activate-cause	0.833	0.882	0.857	0.566	0.625	0.594
ACQ-attribute	0.878	0.935	0.906	0.594	0.92	0.867
SEN-attribute	0.96	0.923	0.941	0.385	0.385	0.385
STP-attribute	0.96	0.827	0.889	1	0.815	0.898

quantitative analysis based on P, R, and F_1, this section compares extracted contents between the proposed AT-NeuroEAE and LDA-based neuroimaging topic modeling by using a specific article. **Figure 7** shows the results of AT-NeuroEAE and LDA for the article titled "Induction of Empathy by the Smell of Anxiety" (Prehn-Kristensen et al., 2009). Different colors represent different categories of entities or topics. The random state was set to 1, the number of topics to 15, and alpha to 0.5.

As shown in **Figure 7**, LDA can obtain multi-aspect topics, which are involved with brain mechanism (e.g., anxiety, frontal gyrus), the experiment (e.g., mind reading task), and analysis (e.g., voxel). However, these topics are isolated and cannot effectively characterize the whole of computational neuroscience research. Some general words, such as "conclusion" and "figure," are also included in topics because a large amount of noise in full-text corpora affects topic recognition based on word distribution.

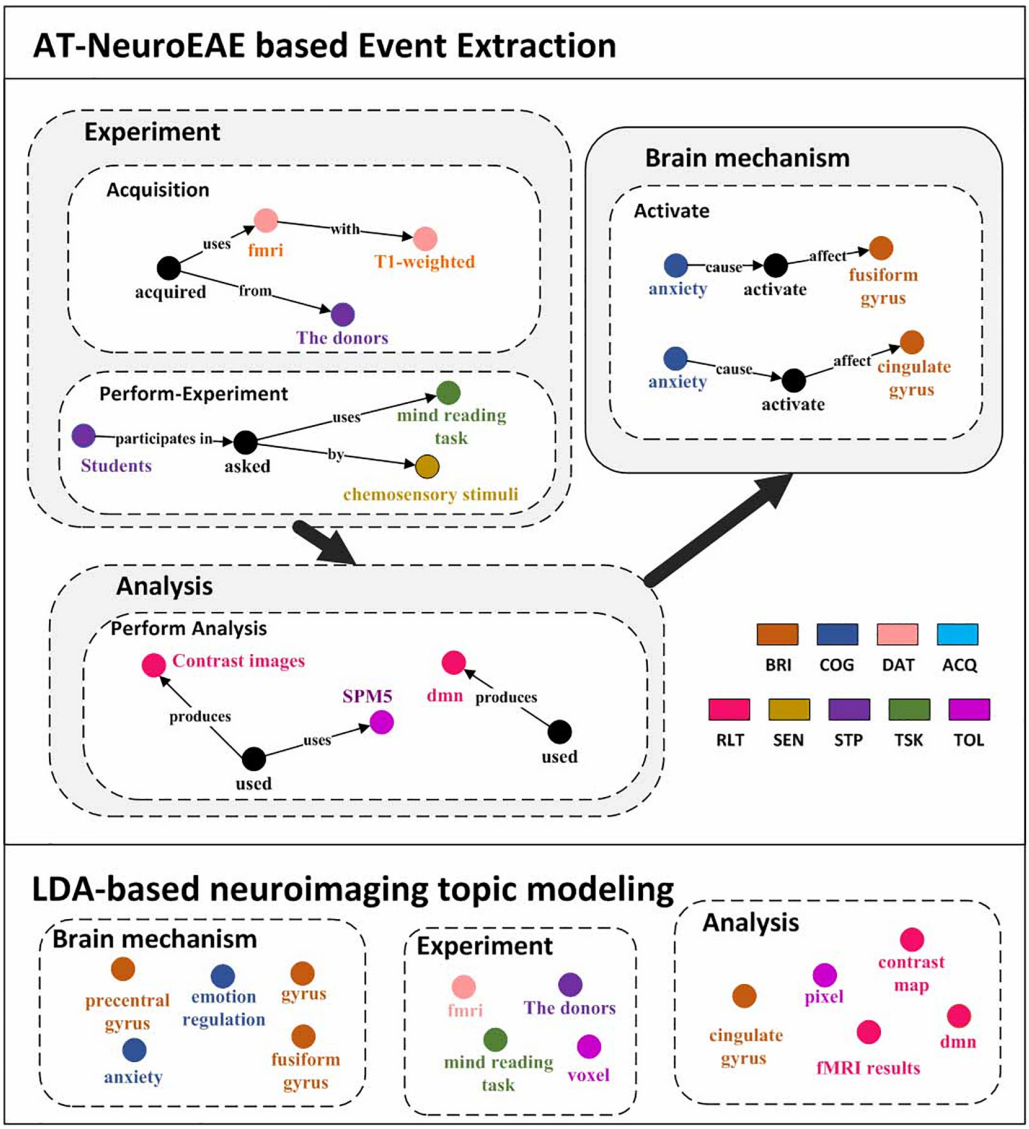


FIGURE 7 | Visualization results of event extraction. The top is five events extracted from the article (Prehn-Kristensen et al., 2009). The bottom is LDA topics extracted from the same article. In order to compare with the results of AT-NeuroEAE, LDA topics are manually divided into three classes, brain mechanism, experiment, and analysis.

Different from LDA, the proposed AT-NeuroEAE can extract a group of events with rich semantics for outlining the whole research process:

- Brain mechanisms: The article reveals that "anxiety" activates the brain areas "cingulate gyrus" and "fusiform gyrus."
- Experiment process: The experiments adopted the "mind reading task" with "chemosensory stimulus." Researchers used "T1 weighted" "fMRI" device to collect data from "the donors."
- Analytical process: The study got the "contrast map" by the "SPM5." "dmn" was also obtained as the analytical result.

Because the extraction is based on pre-defined event categories, those general words can be effectively filtered.

Named entity recognition is the current research focus in neuroimaging text mining. Because related studies (Shardlow et al., 2018; Riedel et al., 2019; Sheng et al., 2019) have different entity or label categories, this study only qualitatively compares their entity/label categories with the proposed AT-NeuroEAE. The result is shown in **Figure 8**.

The study on (Shardlow et al., 2018) mainly focused on brain mechanism, especially multi-level brain structures. The cognitive function and the research process were neglected. In contrast, the study on (Riedel et al., 2019) only took account of the experimental process and neglected brain mechanisms and the analytical process. The study of Sheng et al. (2019) focused on

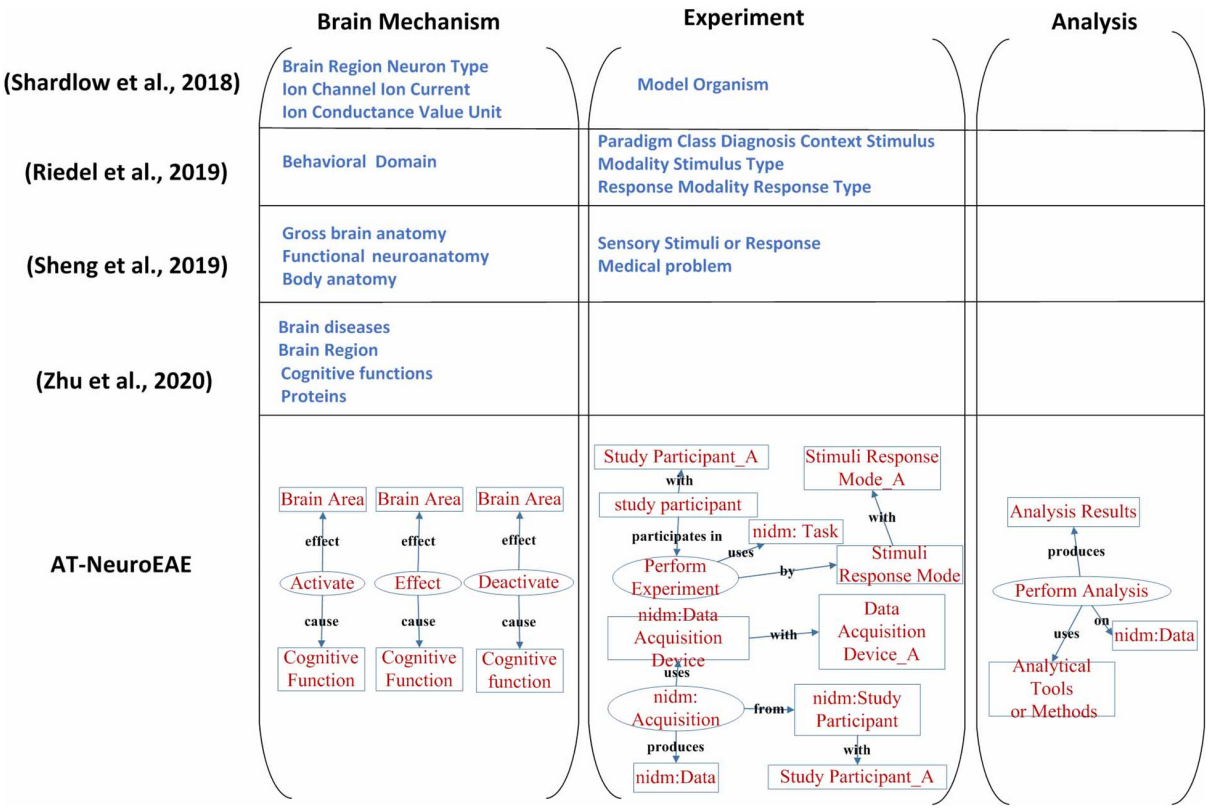


FIGURE 8 | A comparison of neuroimaging entity/label categories/span of interest. The study on Shardlow et al. (2018) mainly focused on brain mechanism, especially multi-level brain structures. The study on Riedel et al. (2019) only took account of the experimental process. The study on Sheng et al. (2019) focused on brain mechanism and two experimental factors, including sensory stimuli or response and study participants' medical problems. The study on Zhu et al. (2020) paid attention on pathology and mechanism of brain diseases. Our study is involved with the whole research process and extracted information is organized by events with rich semantics.

brain mechanism, but neglected the analysis process which is very important to understand and evaluate the conclusion about brain mechanism. About the experiment, only sensory stimuli or response and study participants' medical problems were extracted. Some key experimental factors, including experimental tasks, data acquisition equipment, etc., have not been paid enough attention. The study of Zhu et al. (2020) paid attention on pathology and mechanism of brain diseases, and its spans of interests are limited to brain diseases, brain regions, cognitive functions, and proteins.

In this study, neuroimaging events related to brain mechanism and the whole research process were extracted. In the aspect of brain mechanism, three event categories "Activate," "Deactivate," and "Effect" represent not only location information but also physiological characteristics information such as brain network. In the aspect of experiment, the event category "Perform Experiment" represents as much information as the study on (Riedel et al., 2019) because the representations about stimuli, response, and tasks in neuroimaging articles are flexible and it is not necessary to distinguish the form and type of stimulus, task paradigm, and context in detail. The event category "Acquisition" represents the imaging devices and scanning parameters. In the aspect of analysis, the event category "Perform Analysis" represents three important factors and their relationships during neuroimaging data analysis, including methods/tools, results and brain areas. All extracted information in this study is not an isolated entity, but organized as a systematic information network in the form of events, which can better describe the whole research process.

CONCLUSION

This study proposed an approach based on event extraction to realize research sharing-oriented neuroimaging provenance construction. Guided by the provenance model, the neuroimaging research process and result are modeled as six categories of neuroimaging event-containing attributes. A joint extraction model based on deep adversarial learning, called AT-NeuroEAE, is proposed to extract the defined neuroimaging events in a few-slot learning scenario. The experimental results on the PLOS ONE data set show that the model can realize the large-scale and low-cost neuroimaging provenance construction for open and FAIR research sharing in computational neuroscience.

SOFTWARE DEPENDENCIES

As described above, the analyses presented in this work rely on the following dependencies: numpy (van der Walt et al., 2011), pandas (McKinney, 2011), statsmodels (Seabold and Perktold, 2010), SciPy (Jones et al., 2001), scikit-learn (Pedregosa et al., 2011), IPython (Pérez and Granger, 2007), nltk (Bird et al., 2009), pdfminer (Shinyama, 2007), seaborn (Waskom et al., 2017), and many core libraries provided with Python 2.7.11. Additionally, the ontological expansion of Cognitive Atlas term weights was influenced by Poldrack et al. (2011).

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

SL, JC, and ZX developed the study design and performed the coding and analyses. All authors contributed to the implementation and contributed to the manuscript revision, and read and approved the submitted version.

FUNDING

The work was supported by National Key Research and Development Program of China (Grant No. 2020YFB2104402).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.739535/full#supplementary-material

REFERENCES

Abacha, A. B., Herrera, A., Ke, W., Long, L. R., Antani, S., and Demner-Fushman, D. (2017). "Named entity recognition in functional neuroimaging literature," in *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, 2218–2220. doi: 10.1109/BIBM.2017.8218002

Abera, B. (2020). Event extraction and representation model from news articles. *Int. J. Innov. Eng. Technol.* 16, 1–8.

Abrams, M. B., Bjaalie, J. G., Das, S., Egan, G. F., and Martone, M. E. (2021). A standards organization for open and fair neuroscience: the international neuroinformatics coordinating facility. *Neuroinformatics* 1–12. doi: 10.31219/6osf.io/3rt9b

Agirre, A. G., Marimo, M., Intxaurrondo, A., Rabal, O., and Krallinger, M. (2019). "PharmaCoNER: pharmacological substances, compounds and proteins named entity recognition track," in *Proceedings of the 5th Workshop on BioNLP Open Shared Tasks*, Hong Kong.

Alcalá-López, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogeley, K., Mars, R. B., et al. (2017). Computing the social brain connectome across systems and states. *Cereb. Cortex* 28, 2207–2232. doi: 10.1093/cercor/bhx121

Alhazmi, F. H., Beaton, D., and Abdi, H. (2018). Semantically defined subdomains of functional neuroimaging literature and their corresponding brain regions. *Hum. Brain Mapp.* 39, 2764–2776. doi: 10.1002/hbm.24038

Araki, J., and Mitamura, T. (2018). "Open-domain event detection using distant supervision," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, NM, 878–891.

Armelle, V., Pascale, P., Serge, B., Vincent, D., Béatrice, D., Francis, E., et al. (2014). Episodic future thinking in semantic dementia: a cognitive and fmri study. *PLoS One* 9:e111046. doi: 10.1371/journal.pone.0111046

Arshad, B., Munir, K., Mcclatchey, R., Shamdasani, J., and Khan, Z. (2019). Neuroprov: provenance data visualisation for neuroimaging analyses. *J. Comput. Lang.* 52, 72–87. doi: 10.1016/j.cola.2019.04.004

Badhwar, A. P., Kennedy, D., Poline, J. B., and Toro, R. (2016). Distributed collaboration: the case for the enhancement of Brainspell's interface. *Gigascience* 5, 1–2.

Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., et al. (2015). Using a suite of ontologies for preserving workflow-centric research objects. *J. Web Semant.* 32, 16–42.

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol, CA: O'Reilly Media, Inc.

Bjrne, J., and Salakoski, T. (2018). "Biomedical event extraction using convolutional neural networks and dependency parsing," in *Proceedings of the BioNLP 2018 workshop*, Melbourne, VC.

Bolt, T., Nomi, J. S., Arens, R., Vij, S. G., Riedel, M., Salo, T., et al. (2020). Ontological dimensions of cognitive-neural mappings. *Neuroinformatics* 18, 451–463. doi: 10.1007/s12021-020-09454-y

Buneman, P., Khanna, S., and Wang-Chiew, T. (2004). "Why and where: a characterization of data provenance," in *Proceedings of the International Conference on Database Theory*, (Berlin: Springer). doi: 10.1109/TVCG.2015.2467551

Chen, J. H., and Ning, Z. (2013). Toward the data-brain driven systematic brain data analysis. *IEEE Trans. Syst. Man Cybernet. Syst.* 43, 222–228. doi: 10.1109/TSMCA.2012.2199300

Chen, J. H., Ning, Z., and Liang, P. (2012). Data-brain driven systematic human brain data analysis: a case study in numerical inductive reasoning centric investigation. *Cogn. Syst. Res.* 15–32, 17–32. doi: 10.1016/j.cogsys.2010.12.014

Chen, Y. (2019). Multiple-level biomedical event trigger recognition with transfer learning. *BMC Bioinformatics* 20:459. doi: 10.1186/s12859-019-3030-z

Chung, J. W., Yang, W., and Park, J. (2020). Unsupervised inference of implicit biomedical events using context triggers. *BMC Bioinformatics.* 21:29. doi: 10.1186/s12859-020-3341-0

Curiskis, S. A., Drake, B., Osborn, T. R., and Kennedy, P. J. (2019). An evaluation of document clustering and topic modelling in two online social networks: twitter and reddit. *Inform. Process. Manag.* 57:102034. doi: 10.1016/j.ipm.2019.04.002

Daniel, W., Svenja, T., Anna, B., Münte Thomas, F., Michael, S., Manfred, C., et al. (2014). Tracking functional brain changes in patients with depression under psychodynamic psychotherapy using individualized stimuli. *PLoS One* 9:e109037. doi: 10.1371/journal.pone.0109037

Deléger, L., Bossy, R., Chaix, E., Ba, M., Ferré, A., Bessières, P., et al. (2016). "Overview of the bacteria biotope task at BioNLP shared task 2016," in *Proceedings of the 4th BioNLP Shared Task Workshop*, Berlin, 12–22.

Dinov, I., Van Horn, J., Lozev, K. M., Magsipoc, R., Petrosyan, P., Liu, Z., et al. (2009). Efficient, distributed and interactive neuroimaging data analysis using the loni pipeline. *Front. Neuroinform.* 3:22. doi: 10.3389/neuro.11.022.2009

Doddington, G., Mitchell, A., Przybocki, M. A., Ramshaw, L., Strassel, S., and Weischedel, R. (2004). "The automatic content extraction (ACE) program - tasks, data, and evaluation," in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon. doi: 10.1007/s12021-011-9126-x

Foster, I. (2005). Service-oriented science. *Science* 308, 814–817. doi: 10.1126/science.1110411

French, L., Lane, S., Xu, L., Siu, C., Kwok, C., Chen, Y., et al. (2012). Application and evaluation of automated methods to extract neuroanatomical connectivity statements from free text. *Bioinformatics* 28, 2963–2970. doi: 10.1093/bioinformatics/bts542

Frishkoff, G., Sydes, J., Mueller, K., Frank, R., Curran, T., Connolly, J., et al. (2011). Minimal information for neural electromagnetic ontologies (MINEMO): a standards-compliant method for analysis and integration of eventrelated potentials (ERP) data. *Stand. Genom. Sci.* 5, 211–223. doi: 10.4056/sigs.2025347

Genon, S., Reid, A., Li, H., Fan, L., Müller, V. I., Cieslik, E. C., et al. (2018). The heterogeneity of the left dorsal premotor cortex evidenced by multimodal connectivity-based parcellation and functional characterization. *Neuroimage* 170, 400–411. doi: 10.1016/j.neuroimage.2017.02.034

Gibson, F., Overton, P. G., Smulders, T. V., Schultz, S. R., and Lord, P. (2008). Minimum information about a neuroscience investigation (mini): electrophysiology. *Nat. Preced.* 3. doi: 10.1038/npre.2008.1720.1

Gorgolewski, K. J., Esteban, O., Schaefer, G., Wandell, B. A., and Poldrack, R. A. (2017). OpenNeuro-a free online platform for sharing and analysis of neuroimaging data. *Poster at the 23rd Annual Meeting of the Organization for Human Brain Mapping (OHBM) 2017*, Vancouver, BC, 1677.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwartz, Y., Sochat, V. V., Ghosh, S. S., et al. (2016). NeuroVault.org: a repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. *Neuroimage* 124, 1242–1244. doi: 10.1016/j.neuroimage.2015.04.016

Greenwood, M., Goble, C., Stevens, R., Zhao, J., and Oinn, T. (2003). "Provenance of e-science experiments-experience from bioinformatics," in *Proceedings of The UK OST e-Science Second All Hands Meeting 2003*, Manchester, 223–226.

Huangfu, C., Zeng, Y., and Wang, Y. (2020). Creating neuroscientific knowledge organization system based on word representation and agglomerative clustering algorithm. *Front. Neuroinform.* 14:38. doi: 10.3389/fninf.2020.00038

Jones, E., Oliphant, T., and Peterson, P. (2001). *SciPy: Open Source Scientific Tools for Python*. Available online at: http://www.scipy.org/

Keator, D., Helmer, K., Maumet, C., Padhy, S., Jarecka, D., Ghosh, S., et al. (2019). "Tools for FAIR neuroimaging experiment metadata annotation with NIDM experiment," in *Proceedings of the OHBM 2019 – 25th Annual Meeting of the Organization for Human Brain Mapping Jun 2019*, Rome, 1–5.

Keator, D., Helmer, K., Steffener, J., Turner, J. A., Van Erp, T. G., Gadde, S., et al. (2013). Towards structured sharing of raw and derived neuroimaging data across existing resources. *NeuroImage* 82, 647–661. doi: 10.1016/j.neuroimage.2013.05.094

Kennedy, D. N., Abraham, S. A., Bates, J. F., Crowley, A., Ghosh, S., Gillespie, T., et al. (2019). Everything matters: the reproNim perspective on reproducible neuroimaging. *Front. Neuroinform.* 13:1. doi: 10.3389/fninf.2019.00001

Kip, F. T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv* [Preprint]. arXiv:1609.02907.

Laird, A. R., Lancaster, J. L., and Fox, P. T. (2005). BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* 3, 65–78. doi: 10.1385/ni:3:1:065

Lanter, D. P. (1991). Design of a lineage-based meta-data base for GIS. *Cartogr. Geogr. Inform. Syst.* 18, 255–261. doi: 10.1559/152304091783786718

Lanting, C. P., Kleine, E. D., Langers, D., and Dijk, P. V. (2014). Unilateral tinnitus: changes in connectivity and response lateralization measured with fmri. *PLoS One* 9:e110704. doi: 10.1371/journal.pone.0110704

Lei, B., Wu, F., Zhou, J., Xiong, D., and Wu, K. (2020). Neuro-learn: a solution for collaborative pattern analysis of neuroimaging data. *Neuroinformatics* 19, 79–91. doi: 10.1007/s12021-020-09468-6

Lever, J., and Jones, S. J. (2016). "VERSE: event and relation extraction in the BioNLP 2016 shared task," in *Proceedings of the 4th BioNLP Shared Task Workshop*, Berlin.

Li, D., Huang, L., Ji, H., and Han, J. (2016). "Biomedical event extraction based on knowledge-driven Tree-LSTM," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, vol. 1 (Minneapolis, Minnesota, Hennepin: Association for Computational Linguistics), 1421–1430. doi: 10.18653/v1/N19-1145

Li, L., Liu, Y., and Qin, M. (2020). Extracting biomedical events with parallel multi-pooling convolutional neural networks. *IEEE ACM Trans. Comput. Biol. Bioinform.* 17, 599–607. doi: 10.1109/TCBB.2018.2868078

Ma, X., and Hovy, E. (2016). "End-to-end sequence labeling *via* bi-directional lstm-cnns-crf," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin. doi: 10.13140/RG.2.1.2182.5685

Maumet, C. (2020). "Open Science in action: doing research as a community!," in *Proceedings of the 26th Annual Meeting of the Organization for Human Brain Mapping OHBM 2020*, Montréal, QC, 1–19.

Maumet, C., Auer, T., Bowring, A., Chen, G., Das, S., Flandin, G., et al. (2016). Sharing brain mapping statistical results with the neuroimaging data model. *Sci. Data* 3:160102. doi: 10.1038/sdata.2016.102

Maumet, C., Ghosh, S., Halchenko, Y., Jarecka, D., Nichols, N., Poline, J. B., et al. (2019). "The best of both worlds: using semantic web with JSON-LD. An example with NIDM-Results and Datalad," in *Proceedings of the OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping*, Rome.

McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python High Perform. Sci. Comput.* 14, 1–9.

Miksa, T., and Rauber, A. (2017). Using ontologies for verification and validation of workflow-based experiments. *J. Web Seman.* 43, 25–45. doi: 10.1016/j.websem.2017.01.002

Milham, M. P., Cameron, C. R., Son, J. J., Fleischmann, M., Clucas, J., Xu, H., et al. (2018). Assessment of the impact of shared brain imaging data on the scientific literature. *Nat. Commun.* 9:2818. doi: 10.1038/s41467-018-04976-1

Moreau, L., Ludscher, B., Altintas, I., Barga, R. S., and Zhao, Y. (2008). Special issue: the first provenance challenge. *Concurr. Comput. Pract. Exp.* 20, 409–418.

Mutschler, I., Ball, T., Kirmse, U., Wieckhorst, B., Pluess, M., Klarhöfer, M., et al. (2016). The role of the subgenual anterior cingulate cortex and amygdala in environmental sensitivity to infant crying. *PLoS One* 11:e0161181. doi: 10.1371/journal.pone.0161181

Naud, A., and Usui, S. (2008). Exploration of a collection of documents in neuroscience and extraction of topics by clustering. *Neural Netw.* 21, 1205–1211. doi: 10.1016/j.neunet.2008.05.009

Neumann, N., Lotze, M., and Eickhoff, S. B. (2016). Cognitive expertise: an ale meta-analysis. *Hum. Brain Mapp.* 37, 262–272. doi: 10.1002/hbm.23028

Nguyen, T. M., and Nguyen, T. H. (2019). "One for all: neural joint modeling of entities and events," in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, Honolulu, HI, 6851–6858. doi: 10.1371/journal.pone.0235796

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Pennington, J., Socher, R., and Manning, C. (2014). "Glove: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha.

Pérez, F., and Granger, B. E. (2007). IPython: a system for interactive scientific computing. *Comput. Sci. Eng.* 9, 21–29. doi: 10.1109/MCSE.2007.53

Poldrack, R. A., and Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. *Nat. Neurosci.* 17, 1510–1517. doi: 10.1038/nn.3818

Poldrack, R. A., and Gorgolewski, K. J. (2015). Openfmri: open sharing of task fmri data. *NeuroImage* 144, 259–261. doi: 10.1016/j.neuroimage.2015.05.073

Poldrack, R. A., Aniket, K., Donald, K., Eric, M., Christian, S., Yolanda, G., et al. (2011). The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Front. Neuroinform.* 5:17. doi: 10.3389/fninf.2011.00017

Poldrack, R. A., Fletcher, P. C., Henson, R. N., Worsley, K. J., Brett, M., and Nichols, T. E. (2008). Guidelines for reporting an fMRI study. *Neuroimage* 40, 409–414. doi: 10.1016/j.neuroimage.2007.11.048

Poldrack, R. A., Mumford, J. A., Schonberg, T., Kalar, D., Barman, B., Yarkoni, T., et al. (2013). Discovering relations between mind, brain, and mental disorders using topic mapping. *PLoS Comput. Biol.* 8:e1002707. doi: 10.1371/journal.pcbi.1002707

Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Front. Neuroinform.* 6:9. doi: 10.3389/fninf.2012.00009

Prehn-Kristensen, A., Wiesner, C., Bergmann, T. O., Wolff, S., Jansen, O., Mehdorn, H. M., et al. (2009). Induction of empathy by the smell of anxiety. *PLoS One* 4:e5987. doi: 10.1371/journal.pone.0005987

Riedel, M. C., Salo, T., Hays, J., Md, Turner, and Laird, A. R. (2019). Automated, efficient, and accelerated knowledge modeling of the cognitive neuroimaging

literature using the ATHENA toolkit. *Front. Neurosci.* 13:494. doi: 10.3389/fnins.2019.00494

Rodolphe, N., Allendorfer, J. B., Szaflarski, J. P., and Yoko, H. (2014). A model for visual memory encoding. *PLoS One* 9:e107761. doi: 10.1371/journal.pone.0107761

Ruiz-Olazar, M., Rocha, E. S., Rabaa, S. S., Ribas, C. E., and Braghetto, K. R. (2016). "A review of guidelines and models for representation of provenance information from neuroscience experiments," in *Proceedings of the 6th International Workshop on Provenance and Annotation of Data and Processes*, (Berlin: Springer-Verlag).

Seabold, S., and Perktold, J. (2010). "Statsmodels: econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference*, Vol. 57. (San Francisco, CA: Scipy), 61.

Shafahi, A., Najibi, M., Ghiasi, A., Xu, Z., Dickerson, J., Studer, C., et al. (2019). "Adversarial training for free!," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vol. 302, (Red Hook, NY: Curran Associates Inc.), 3358–3369.

Shardlow, M., Ju, M., Li, M., O'Reilly, C., Iavarone, E., McNaught, J., et al. (2018). A text mining pipeline using active and deep learning aimed at curating information in computational neuroscience. *Neuroinformatics* 17, 391–406. doi: 10.1007/s12021-018-9404-y

Sheng, Y., Chen, J., He, X., Xu, Z., and Lin, S. (2020). A topic learning pipeline for curating brain cognitive researches. *IEEE Access* 8, 191758–191774. doi: 10.1109/ACCESS.2020.3032173

Sheng, Y., Lin, S., Gao, J., He, X., and Chen, J. (2019). "Research sharing-oriented functional neuroimaging named entity recognition," in *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Vol. 2019, (San Diego, CA: IEEE Press), 1629–1632.

Shinyama, Y. (2007). *PDFMine-Python PDF Parser*. Available online at: http://pypi.python.org/pypi/pdfminer/ (accessed April 29, 2016).

Simmhan, Y. L., Plale, B., and Gannon, D. (2005). A survey of data provenance in e-science. *ACM SIGMOD Record.* 34, 31–36. doi: 10.1145/1084805.1084812

Soomro, P. D., Kumar, S., Banbhrani, Shaikh, A. A., and Raj, H. (2017). Bio-NER: biomedical named entity recognition using rule-based and statistical learners. *Int. J. Adv. Comput. Sci. Appl.* 8, 163–170. doi: 10.14569/IJACSA.2017.081220

Sun, R., Guo, S., and Ji, D. H. (2017). Topic representation integrated with event knowledge. *Chin. J. Comput.* 40, 791–804. doi: 10.11897/SP.J.1016.2017.00791

Task Force on Grand Challenges (2011). *NSF Advisory Committee for Cyberinfrastructure*. Arlington, VA: National Science Foundation.

That, D., Fils, G., Yuan, Z., and Malik, T. (2017). Sciunits: reusable research objects. *arXiv* [Preprint]. arXiv:1707.05731v2,

Trieu, H. L., Tran, T. T., Duong, K. N. A., Nguyen, A., Miwa, M., and Ananiadou, S. (2020). DeepEventMine: end-to-end neural nested event extraction from biomedical texts. *Bioinformatics* 36, 4910–4917. doi: 10.1093/bioinformatics/btaa540

Turner, J. A., and Laird, A. R. (2012). The cognitive paradigm ontology: design and application. *Neuroinformatics* 10, 57–66. doi: 10.1007/s12021-011-9126-x

van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22. doi: 10.1109/MCSE.2011.37

Van Horn, J. D., Grethe, J. S., Kostelec, P., Woodward, J. B., Aslam, J. A., Rus, D., et al. (2001). The functional magnetic resonance imaging data center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 1412, 1323–1339. doi: 10.1098/rstb.2001.0916

Wang, G., Li, C., Wang, W., Zhang, Y., and Carin, L. (2018). "Joint embedding of words and labels for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, VC.

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., et al. (2017). *mwaskom/seaborn: v0.8.1 (Version v0.8.1)*. doi: 10.5281/zenodo.883859

Wegrzyn, M., Aust, J., Barnstorf, L., Gippert, M., Harms, M., Hautum, A., et al. (2018). Thought experiment: decoding cognitive processes from the fmri data of one individual. *PLoS One* 13:e0204338. doi: 10.1371/journal.pone.0204338

Xiang, W., and Wang, B. (2019). A survey of event extraction from text. *IEEE Access* 7, 173111–173137. doi: 10.1109/ACCESS.2019.2956831

Xu, K., Yang, Z., Kang, P., Wang, Q., and Liu, W. (2019). Document-level attention-based bilstm-crf incorporating disease dictionary for disease named entity recognition. *Comput. Biol. Med.* 108, 122–132. doi: 10.1016/j.compbiomed.2019.04.002

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635

Yu, X., Rong, W., Liu, J., Zhou, D., Ouyang, Y., and Xiong, Z. (2019). LSTM-based end-to-end framework for biomedical event extraction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17, 2029–2039. doi: 10.1109/TCBB.2019.2916346

Yuan, Z., Ton That, D. H., Kothari, S., Fils, G., and Malik, T. (2018). Utilizing provenance in reusable research objects. *Informatics* 5:14.

Zhang, S., Sheng, Y., Gao, J., Chen, J., Huang, J., and Lin, S. (2019). "A multi-domain named entity recognition method based on part-of-speech attention mechanism," in *Proceedings of the CCF Conference on Computer Supported Cooperative Work and Social Computing* (Singapore: Springer), 4.

Zhao, W. Z., Zhao, Y., Jiang, X. P., He, T., Liu, F., and Li, N. (2021). Efficient multiple biomedical events extraction *via* reinforcement learning. *Bioinformatics* 37, 1891–1899. doi: 10.1093/bioinformatics/btab024

Zheng, S., Cao, W., Xu, W., and Bian, J. (2019). "Doc2edag: an end-to-end document-level framework for chinese financial event extraction," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong. doi: 10.18653/v1/D19-1032

Zheng, S., Hao, Y., Lu, D., Bao, H., Xu, J., Hao, H., et al. (2017). Joint entity andrelation extraction based on a hybrid neural network. *Neurocomputing* 257, 59–66. doi: 10.1016/j.neucom.2016.12.075

Zhu, H., Zeng, Y., Wang, D., and Huangfu, C. (2020). Species classification for neuroscience literature based on span of interest using sequence-to-sequence learning model. *Front. Hum. Neurosci.* 14:128. doi: 10.3389/fnhum.2020.00128

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

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fnins.2022.832824

Residual Block Based Nested U-Type Architecture for Multi-Modal Brain Tumor Image Segmentation

Sirui Chen[1], *Shengjie Zhao*[1]* *and Quan Lan*[2]*

[1] School of Software Engineering, Tongji University, Shanghai, China, [2] Department of Neurology, First Affiliated Hospital of Xiamen University, Xiamen, China

Multi-modal magnetic resonance imaging (MRI) segmentation of brain tumors is a hot topic in brain tumor processing research in recent years, which can make full use of the feature information of different modalities in MRI images, so that tumors can be segmented more effectively. In this article, convolutional neural networks (CNN) is used as a tool to improve the efficiency and effectiveness of segmentation. Based on this, Dense-ResUNet, a multi-modal MRI image segmentation model for brain tumors is created. The Dense-ResUNet consists of a series of nested dense convolutional blocks and a U-Net shaped model with residual connections. The nested dense convolutional blocks can bridge the semantic disparity between the feature maps of the encoder and decoder before fusion and make full use of different levels of features. The residual blocks and skip connection can extract pixel information from the image and skip the link to solve the traditional deep traditional CNN network problem. The experiment results show that our Dense-ResUNet can effectively help to extract the brain tumor and has great clinical research and application value.

Keywords: brain tumor, multi-modal image segmentation, MRI, CNN, ResNet, UNet

OPEN ACCESS

Edited by:
Yizhang Jiang,
Jiangnan University, China

Reviewed by:
Jing Xue,
Wuxi People's Hospital Affiliated to Nanjing Medical University, China
Zhifan Gao,
Sun Yat-sen University, China

***Correspondence:**
Shengjie Zhao
shengjiezhao@tongji.edu.cn
Quan Lan
xmdylanquan@163.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *10 December 2021*
Accepted: *03 February 2022*
Published: *09 March 2022*

Citation:
Chen S, Zhao S and Lan Q (2022) Residual Block Based Nested U-Type Architecture for Multi-Modal Brain Tumor Image Segmentation. Front. Neurosci. 16:832824. doi: 10.3389/fnins.2022.832824

1. INTRODUCTION

Tumors are localized cell growths in the body that constitute a tumor for various reasons. There are two classifications of tumors, benign and malignant. Cancer is usually a malignant tumor, and is a common malignant diseases that threaten human health. Therefore, accurate segmentation and subsequent quantitative analysis of tumor images is a routine and critical step in treatment.

MRI is based on the principle that the electromagnetic signal released by the instrument has different attenuation in different structures in the body to obtain the electromagnetic signal from the body and reconstruct the body information (Vaughan et al., 2006). Compared with other traditional medical imaging techniques such as X-ray and CT imaging, MRI can provide early detection of smaller and more microscopic lesions (Nandpuru et al., 2014). Due to the complexity of human organ tissues, the same imaging technique usually results in images of different modalities, where different modalities also reveal different pathological information (Legg et al., 2015). MRI provides four different imaging sequences obtained by several different imaging display techniques, and these four different imaging sequences can provide complementary information for clinical diagnosis.

At present, the segmentation of MRI brain tumor images is mainly based on the experience of expert doctors, but it is difficult to accurately segment MRI brain tumor images preoperatively

because of the time-consuming and repetitive work, and the subjective judgment of the physician can interfere with it. In recent years, many researchers have improved natural image segmentation algorithms and then migrated them to brain tumor segmentation tasks, resulting in a large number of effective research methods and findings. As shown in **Figure 1**, we can classify them into two categories: (i) Fully automated segmentation methods based on deep learning and (ii) Traditional semi-automatic medical image segmentation methods.

Edge-based detection: The basic idea of the edge detection-based image segmentation method is to first find its edge pixels in the image and then aggregate these pixels to form the desired region edges and segment the image with the region edges. Aslam et al. (2015) proposed an improved edge detection algorithm for brain tumor segmentation based on the operator (Sobel, 1970) of the first-order derivatives, which combined the Sobel method and the image dependent thresholding method and used a closed contour algorithm to find the different regions. Finally, the intensity information within the closed contours was used to extract the tumor from the image.

Threshold-based: The idea of the algorithm is that first, the grayscale features of the image are computed to obtain one or more grayscale thresholds, and then the grayscale values of each pixel in the image are compared with each other and with the computed thresholds. Finally, based on the comparison results, the pixels can be classified into different classes. Kittler and Illingworth (1986) proposed a computationally efficient solution to the minimum error thresholding problem applicable to multiple threshold selection. Saad et al. (2010) used the histogram thresholding technique in segmentation to detect pixels with high or low intensity.

Region-based: Region-based image segmentation algorithms can be divided into two basic forms: (i) The global departure form, where segmentation is performed gradually until the desired region and (ii) The region growth form, where individual image pixels are started and gradually merged, which in turn forms the segmented region we need. Kaus et al. (2001) used a region-growing approach to segment MRI brain tumor images based on signal intensity values.

Since the main idea of traditional segmentation algorithms is to start from features such as texture and grayscale values between different tissues in medical images, the main challenges of this type of algorithms are the large grayscale similarity and uneven distribution between brain tissues in MR images, while the grayscale values contain too little information and the variability between different cases, which can affect the final segmentation accuracy.

Machine learning-based: Machine learning-based image segmentation algorithms require training a model from a certain number of images with well-defined mapping relationships between image features and labels to learn the segmentation laws. Weakly supervised learning methods and semi-supervised learning methods were first developed, mainly random forests (Shi and Horvath, 2005), Adaboost (Rtsch et al., 2001), K-Means clustering (Hartigan and Wong, 2013), support vector machines (SVMs) (Zhang et al., 2006), etc. The method proposed by Abdelmaksoud et al. (2015) can not only take advantage of the fuzzy C-means algorithm to obtain high accuracy, but also make full use of K-Means clustering to obtain the minimum computation time, which decreased the segmentation time while improving the segmentation accuracy. Tustison et al. (2014) combined the random forest model with regularized probability and used the probability map generated by this model for Markov random field, which finally obtained better results for brain tumor segmentation. Mahmood and Basit (2016) proposed an automatic segmentation framework for ischemic stroke lesion segmentation in multi-spectral MRI based on random forest. However, the problem with the machine learning-based approach is that the selection and labeling of image features require the use of specialized medical knowledge, which limits the development of such methods.

CNNs were originally used in areas such as handwritten alphabet classification, but the input of such images was already fixed in size, so when migrating such networks to the field of image segmentation, it also started by classifying blocks of

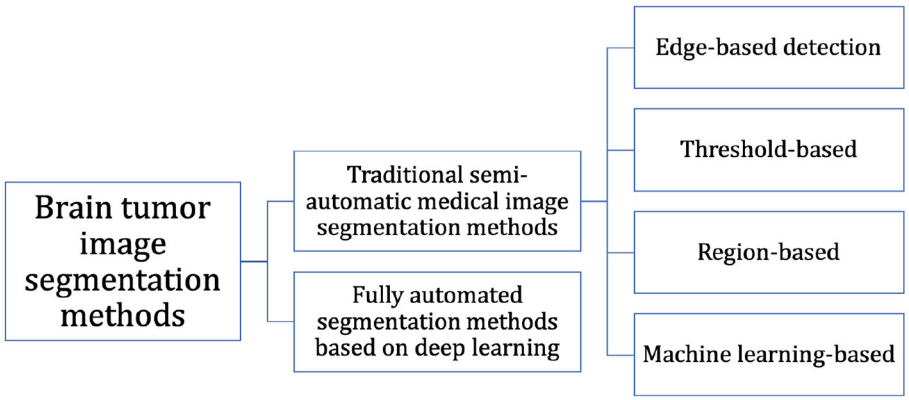


FIGURE 1 | Classification of brain tumor image segmentation methods.

images of fixed size. Moeskops et al. (2016) came up with a method to automatically segment MRI brain images. Their network obtained multi-scale information for each voxel, using multiple convolutional kernels of different sizes. The method did not rely on explicit features but learns to identify important information about the classification based on the training data. However, this method was used only for a single modality MRI image. Hou et al. (2016) proposed a segmentation network that cut a complete image into fixed-size image blocks, after which they were convolved separately to extract features and finally classified for the central pixel points to obtain the final segmentation results. Their method focused the classification from the image level to the pixel level. Brosch (2015) proposed a 3D deep neural network, this model had both convolutional and deconvolutional layers and combined feature extraction and segmentation prediction, this method has high accuracy but was too computationally intensive.

The high cost of acquiring medical images and the relatively small dataset for medical image segmentation also pose a problem for deep learning that requires large amounts of data. And due to the layer-by-layer convolution of neural networks, the accuracy of segmentation is reduced while increasing the computational effort. Multi-modal MRI brain tumor image segmentation can make full use of the feature information of different modalities in MRI thus improve the effectiveness of segmentation, which is one of the recent research hotspots in the field of brain tumor image processing. Using artificial intelligence technology, a fully automated method for brain tumor image segmentation can be designed so that it can support physicians' analysis and diagnosis, which is the best way to solve the above problem.

In general, this article focuses attention on how to fully and efficiently utilize multi-modal MRI image information, proposes a multi-modal nested dense ResU-Net segmentation network, and validates the effectiveness of the algorithm on an open-source dataset of brain tumors.

2. MATERIALS AND METHODS

The model proposed in this section, given the abbreviation Dense-ResUNet, is a multi-modal MRI brain tumor segmentation model based on the traditional U-Net (Ronneberger et al., 2015) and U-Net++ (Zhou et al., 2018), and the network structure is shown in **Figure 2**. The left is the downsampling path, which effectively extracts the image features through a series of successive convolution and downsampling operations, and reduces the image size while increasing the number of channels; the right is the upsampling path, which successfully recovers the image size, improves the image segmentation accuracy and allows better reconstruction of details through a series of successive transposition convolution operations; the middle is a series of nested dense convolutional blocks. The structure bridges the semantic disparity between the feature maps of the encoder and decoder before fusion. Based on this modular design, our model can train the network by adding a small number of convolutional layers to extract good medical image features.

The shalllow convolution structure cannot fully capture the complex structure of the image, but the deep convolution and redundant structure of the stack lead to the gradient vanishing and tearing problem. Therefore, we use a residual unit to extract pixel information from the image and skip the link to solve the traditional deep CNN network problem. The structure of the residual unit is shown in **Figure 3** and the specific embedding method is shown in **Figure 5**. Each layer in the constructed Dense-ResUNet model needs to use convolutional blocks so that feature extraction can be performed. Each of these convolutional blocks consists of 2 convolutional units. In each of these 2 convolutional units, a BatchNorm (BN) layer is included to improve model convergence. The activation is then performed with the ReLU function, which is used to improve the non-linearity of the function.

2.1. ResNet Architecture Overview

Each of the convolutional layers contains feature maps. When a pixel in the image is scanned by the convolution kernel and makes full use of the complex content information in the environment to generate semantic image features, the image features with content and space are expressed by the activation function (ReLU) as Equation (1):

$$X_j^{l+1} = f(t_i^{l+1} + \sum_{i \in I_j} X_i^l \divideontimes k_{ij}^{l+1}) \tag{1}$$

Where $X_j{}^{l+1}$ represents the feature map after the $(l+1)-th$ layer, t represents the offset term, $X_i{}^l$ represents the input feature in the $(l+1)-th$ layer, f is the activation function (rectifier linear unit, ReLU), I_j represents a series of input eigenmatrices, $\divideontimes$ represents the convolution operation, k represents the convolution kernel.

By reducing the dimension of the image features, the pooling layer can represent the high-level content and semantics:

$$X_j^{l+1} = t_j^{l+1} + X_j^l \circledast k_j^{l+1} \tag{2}$$

Where $\circledast$ refers to pooling operations in the convolutional structure. Finally, the fully connected layer takes the maximum likelihood function as the prediction layer to carry out the image classification task.

Figure 4 illustrates how a building block works in ResNet. A quick connection mechanism between each initial input X and the module output $H(X)$ is introduced, so that the input can learn the residual expression $F(X) = H(X) - x$ directly to model the target output $[F(X) + X]$. With such a mechanism, the precision problem and performance degradation due to too many stacked convolution structures can be avoided, and such a quick connection mechanism can perform identity mapping in multi-layer structures. It serves as a reference for the input elements in each layer and learns to form the corresponding residual function instead of some function blocks with no practical value $[F'(X)]$. It is easier to optimize the propagation using such a mapping method and thus can significantly increase the number of layers

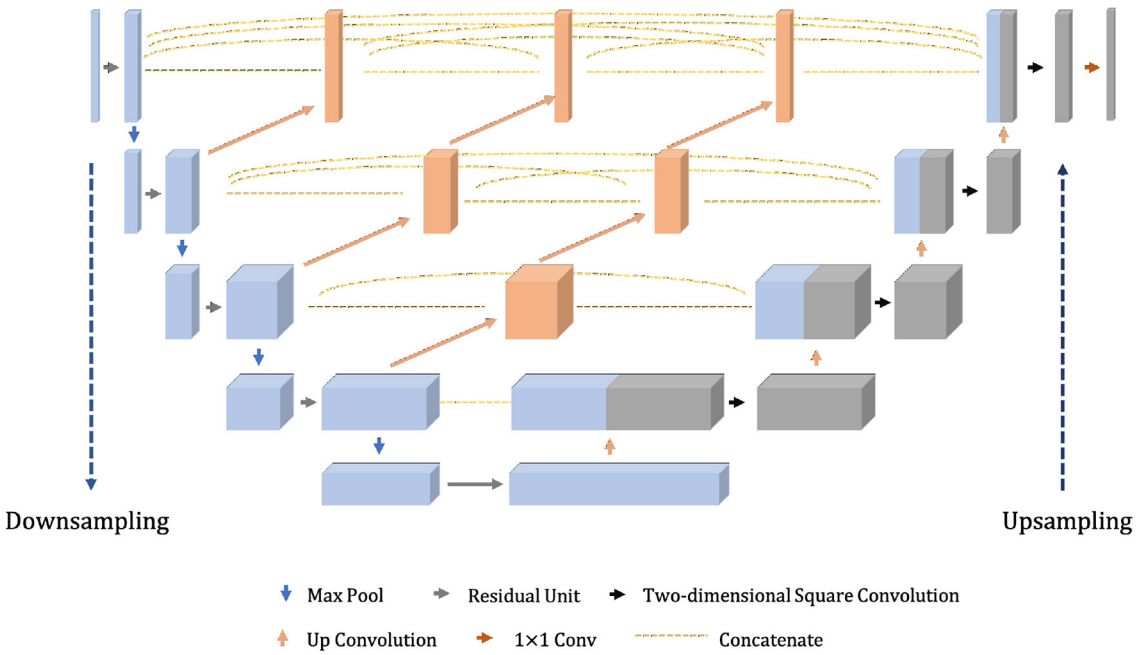


FIGURE 2 | The overall structure of our proposed Dense-ResUNet.

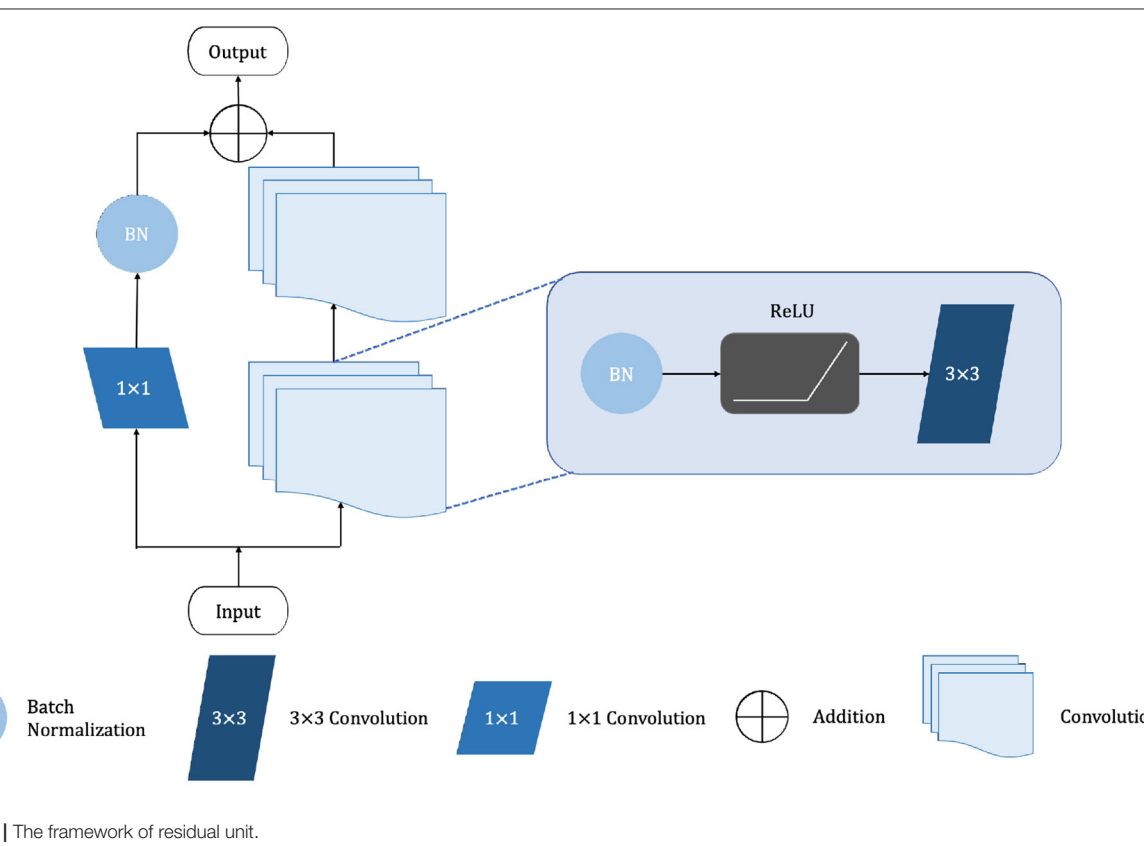


FIGURE 3 | The framework of residual unit.

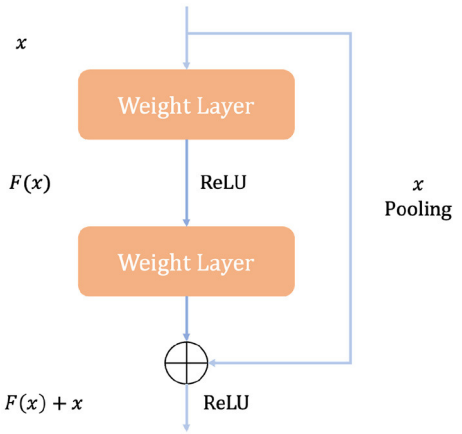


FIGURE 4 | The basic block of ResNet.

in the network structure. The formula for the residual mapping function is as follows:

$$F = W_2 f(W_1 x) \tag{3}$$

$$y = F(x, W_i) + W_s x \tag{4}$$

Where $f(\cdot)$ represents the ReLu activation function and then generates the corresponding output y through a Shortcut Connection and the second ReLU function. When we switch the dimensions of the input and output representations, we can use W_s for linear transformation.

2.2. UNet Architecture Overview

Medical images usually have small samples and most medical images have complex modalities, and the U-Net model proposed by Ronneberger et al. (2015) has an excellent performance in processing medical images with small samples due to its excellent learning ability. The main idea of U-Net is to connect a network with a similar structure to the previous layer behind the downsampling stage, and then recover the resolution of the image output by applying upsampling and combining the output of upsampling with downsampling having high-resolution features combined in these ways to better extract the edge features of the image. **Figure 5** demonstrates the network framework.

Suppose F_{mn} denotes the convolution operation performed at position (m, n), the size of the two-dimensional convolution kernel is $a \times b$, and the two-dimensional convolution is calculated as Equation (5).

$$F_{mn} = f(s_{mn} + \sum_{i=0}^{a-1} \sum_{j=0}^{b-1} w_{ij} \cdot X_{(m+i)(n+j)}) \tag{5}$$

Where s refers to stride, w refers to convolution kernel and X is the input.

From **Figure 2**, we can see that the feature maps of the encoder go through a dense convolution block. The convolution layers' number of the blocks depends on the pyramid level. We assume that X^{l_d,l_c} is a node in the model, where l_d refers to the downsampling layer along the encoder and l_c refers to the convolution layer of the dense block along the skip connection. Meanwhile, we define x^{l_d,l_c} as the output of X^{l_d,l_c}, then the x^{l_d,l_c} can represent the feature maps as Equation (6).

$$x^{l_d,l_c} = \begin{cases} \Delta(x^{l_d-1,l_c}), & where \quad l_c = 0 \\ \Delta([[x^{l_d,l_k}]_{l_k=0}^{l_c-1}, \mathcal{F}(x^{l_d+1,l_c-1})]), & where \quad l_c > 0 \end{cases} \tag{6}$$

Where $\Delta(\cdot)$ refers to the convolution operation followed by ReLU, $\mathcal{F}(\cdot)$ refers to the upsampling operation and [] refers to the concatenate operation.

The role of the pooling layer is to perform merge operations on the input data. We conducted the maximum pooling approach in this article. We calculate the height and weight of the output according to Equations (7) and (8).

$$H_{out} = \lfloor \frac{H_{in} + 2 \times p_i - d_i \times (k_i - 1) - 1}{s_i} + 1 \rfloor \tag{7}$$

$$W_{out} = \lfloor \frac{W_{in} + 2 \times p_j - d_j \times (k_j - 1) - 1}{s_j} + 1 \rfloor \tag{8}$$

Where p refers to the padding, d refers to the dilation, H refers to height and W refers to weight.

The objective function used in the network proposed in this article is dice loss, which is a concept first proposed by Milletari et al. (2016). Its functional expression is shown in Equation (9) and the function takes values in the range [0,1].

$$D = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{9}$$

where p_i is the pixel value of point i in the prediction result and g_i is the pixel value of point i in the label, resulting in the gradient equation as in Equation (10).

$$\frac{\partial D}{\partial p_i} = 2[\frac{g_i \sum_i^N p_i^2 + \sum_i^N g_i^2 - 2p_i \sum_i^N p_i g_i}{(\sum_i^N p_i^2 + \sum_i^N g_i^2)^2}] \tag{10}$$

2.3. Multi-Modal MRI

MRI is performed by varying the direction and intensity of the magnetic field thus obtaining different imaging sequences, which is called multi-modal. Four main imaging forms of multi-modal usually exist, namely: T1-weighted form, T1c-weighted form, T2-weighted form, and FLAIR-weighted form. The four modalities are shown in order from left to right in **Figure 6**, and these pictures also show that the brain tumor image has complex texture and obvious structural information. The imaging of different modalities can highlight different characteristic information of the tumor. By deeply analyzing and thinking about multi-modal images, we can obtain more comprehensive information of brain tumor images and provide help to the research related to brain tumor lesion regions.

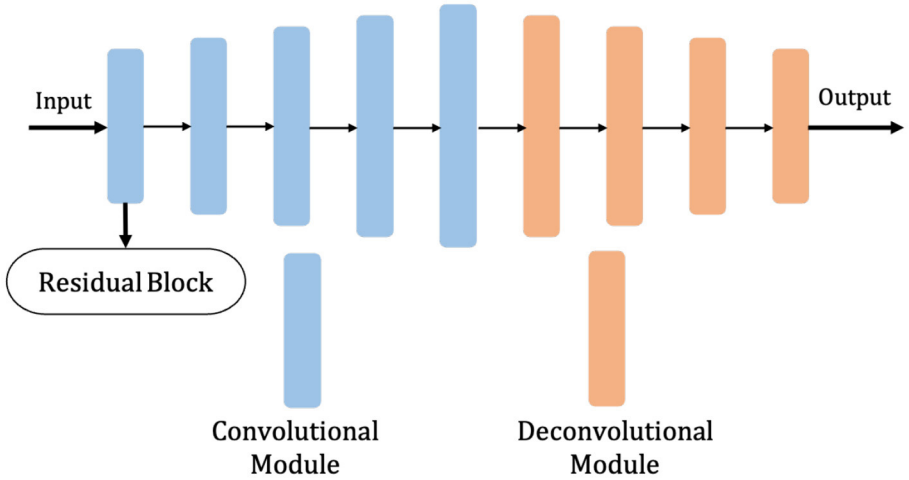


FIGURE 5 | Dense-ResUNet model framework diagram.

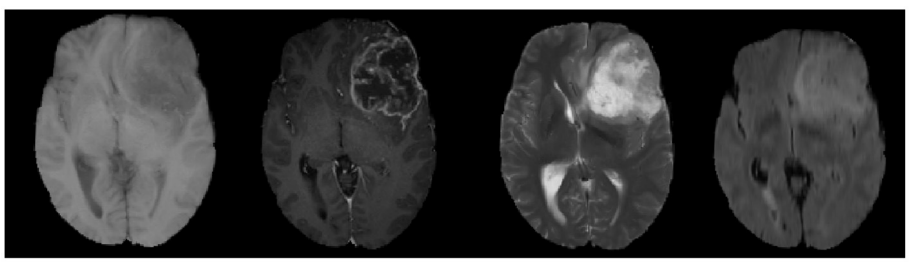


FIGURE 6 | MRI images of four different modalities.

2.4. Datasets

We use BraTS 2018 dataset to train the modal, and it consists of high-grade gliomas (HGG), low-grade gliomas (LGG), and real known tumor segments repeatedly hand-drawn by several medical experts. Four MRI sequences were performed in each case. All imaging sequences of the same subject were preprocessed and stripped from the skull, and all 3D MRI images were volume sized 240 × 240 × 155 and they were aligned, interpolated, and rescaled to achieve the same resolution ($1mm^3$). A total of five types of intra-tumor structures were available in the dataset: (i) normal tissue; (ii) edematous regions (Peritumoral Edema, ED); (iii) necrosis (NCR); (iv) Non-enhancing Tumor (NET); and (v) Enhancing Tumor (ET).

2.5. Multi-Modal Fusion of Brain Tumor Images

Multi-modal brain tumor image fusion is the synthesis of multiple images into a new image. The fusion result is more beneficial to human recognition and automatic machine detection due to the ability to exploit the spatio-temporal correlation and information complementarity of multiple brain tumor images, and to make the fused image obtained a more comprehensive and clear description of the scene.

Our proposed method performs pixel-level image fusion of images and constructs multi-modal image channels, as show in Equation (11). The image obtained after pixel-level image fusion preserves the original information to the maximum extent, which is conducive to further analysis, processing, and understanding of the image, and is also able to expose potential targets, facilitating the operation of judgment to identify potential target pixel points, which is the only way to preserve as much information as possible in the source image, making the fused image increased in both content and detail, an advantage that is unique and exists only in pixel-level fusion.

$$F_{COMB}^{(i)} = F_{Flair}^{(i)} + F_{T1}^{(i)} + F_{T1c}^{(i)} + F_{T2}^{(i)} \tag{11}$$

3. EXPERIMENTAL CONFIGURATIONS

3.1. Evaluation Metrics

In this article, the evaluation metrics proposed by MICCAI BraTS are used to evaluate the performance of the proposed segmentation method. The dataset MICCAI BraTS provides five types of labels for training the network. Different structures in the five types of labels are divided into three regions to meet the medical clinical applications in real situations, and the evaluation

method described in this chapter also focuses on these three tumor regions.

(i) **Enhancing tumor (ET)**, which is present only in high-grade gliomas and is a characteristic enhancing core structure; (ii) **Tumor core (TC)**, a collection of three types of tumors: necrotic, non-enhancing and enhancing tumors; (iii) **Whole-area tumor (WT)**, which is a collection of all intra-tumor structures.

T_1 denotes the real brain tumor area (Ground Truth). T_0 denotes other parts, i.e., normal brain regions. P_1 denotes the predicted brain tumor area, P_0 denotes the other parts are the predicted normal brain regions. It is assumed that brain tumor is a positive sample and normal brain tissue is a negative sample. In the actual evaluation, (i) TP: the part of brain tumor correctly segmented by the model; (ii) TN: the part of normal tissue correctly segmented by the model; (iii) FP: the part of normal tissue but segmented as brain tumor; and (iv) FN: the part of brain tumor but segmented as normal tissue, which have also been shown in **Table 1**. Thus, we utilize the following evaluation metrics to evaluate our image classification model:

(i) Dice Similarity Coefficient (DSC) is a kind of set similarity measure, usually used to calculate the similarity of two samples, the higher the Dice coefficient means that the segmentation result of the algorithm is closer to the real result.

$$Dice(P,T) = \frac{|P_1 \cap T_1|}{\frac{(|P_1|+|T_2|)}{2}} = \frac{2TP}{FP+2TP+FN} \tag{12}$$

(ii) Sensitivity indicates the intersection of the results of the segmentation algorithm and the real results than the value of the real results, the larger the value of this evaluation index means that the segmentation results are closer to the real results.

$$Sensitivity = \frac{|P_1 \cap T_1|}{|T_1|} = \frac{TP}{TP+FN} \tag{13}$$

(iii) Positive Predictive Value (PPV) refers to the proportion of cases that are tumors among the number of positive samples of tumor segmentation results to be evaluated.

$$PPV = \frac{TP}{TP+FP} \tag{14}$$

(iv) Hausdorff distance, which represents the maximum distance between the edge points segmented by the algorithm and the actual edge points.

$$Hausdorff(P,T) = \max\{h(P,T), h(T,P)\} \tag{15}$$

3.2. Data Pre-processing

The dataset has completed the basic steps of brain MRI image processing (image alignment, cranial separation, etc.), but due to the specificity of medical images, there are some differences in MRI images obtained from the same patient machine trial in the same scanner at different time points. Considering the variability of MRI brain tumor image data distribution, to make the contrast and intensity range of patient MRI images similar, this section standardizes and normalizes the pixel intensity to balance the difference of gray value between images, and the process is as follows (also shown in **Figure 7**):

(1) Removal of 1% of the lowest and highest intensity pixel values.

(2) For each case, the mean value of the image is subtracted and divided by the standard deviation of the image, and data normalization is performed on the data within each input channel to obtain the standardized image.

$$F = \frac{X-\mu}{std} \tag{16}$$

In Equation (16), the X denotes the image matrix, and μ represents the pixel-average value of the image.

$$std = \max(\sigma, \frac{1}{\sqrt{N}}) \tag{17}$$

TABLE 1 | Pixel point classification labels.

Classification A[a] \ Classification B[b]	Brain tumor area	Normal tissue area
High-grade glioma	TP	FN
Tumor Tag	FP	TN

[a] *Classification of pixel dots in the standard.*
[b] *Classification of pixel points in prediction.*

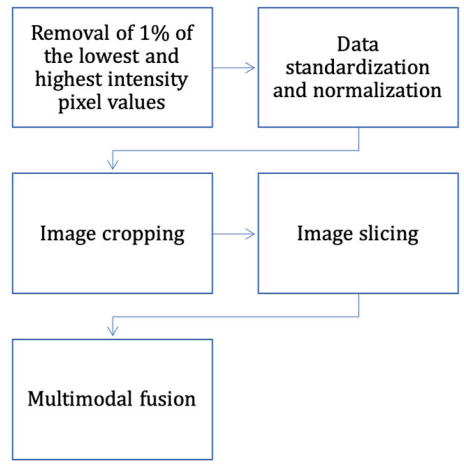


FIGURE 7 | The process of data pre-processing.

TABLE 2 | Training parameters setting.

Parameter	Value
Learning rate	0.0003
Batch_size	10
Early_stop	20
Epochs	1000
Optimizer	Adam

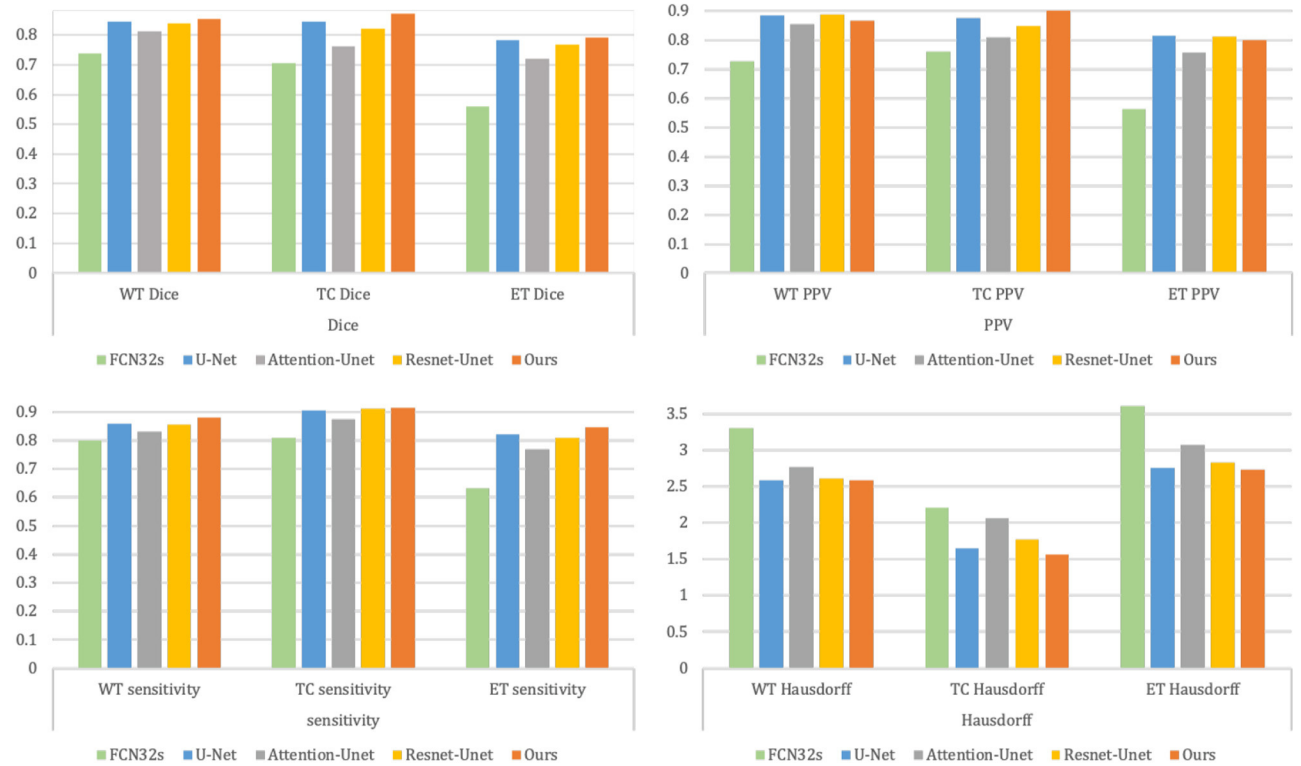


FIGURE 8 | Results comparison between the baseline methods and our proposed methodology.

In Equation (17), the σ denotes the standard variance of the image, and N denotes the total number of pixels of the matrix X represented by the image.

(3) Normalization operation is performed for image slices to compress their pixel intensities.

The intensity of the MRI images of the brain after the three processing steps described above is in the range of [0, 1]. Most medical images are 3D data, but the expert annotations in the BraTS dataset are for 2D axial slices rather than 3D images. In this article, two-dimensional brain tumor images were used for analysis as using 2D slices as network input can also provide enough features to identify each tumor region; therefore, the only way to fit the two-dimensional network was to slice the three-dimensional data into two-dimensional data. The problem of segmenting brain tumors in MRI images is associated with a large data imbalance. To reduce the imbalance of categories, we crop the image slices from the axial plane extraction into 2D image blocks of 128 × 128. The cropped sampled image blocks not only well avoid the zero-intensity pixel blocks, but also reduce the size of normal tissue regions, which are helpful to alleviate the data imbalance problem. In addition, this also reduces the computational effort and the training time.

3.3. Implementation Details

The Pytorch framework is used to implement the model and the experiments are done on Ubuntu 20.04 system with an Intel Core i7 3.5GHz processor and an NVIDIA TITAN 2080Ti graphics card. Eighty percent (i.e., 233 patients) was used to train the model There are four hyperparameters set for Adam in Pytorch: lr (learning rate), smoothing constant betas, eps and weight_decay. The parameters of our proposed model are shown in **Table 2**.

TABLE 3 | Results comparison.

Evaluation metrics	FCN32s	UNet	Attention-UNet	ResNet-UNet	Ours
WT Dice	0.7376	0.8450	0.813	0.8384	**0.8529**
TC Dice	0.7070	0.8454	0.7633	0.8228	**0.8705**
ET Dice	0.5610	0.7817	0.7218	0.7685	**0.7908**
WT PPV	0.7287	0.8859	0.8567	0.8887	0.8686
TC PPV	0.7599	0.8775	0.8105	0.8478	**0.9050**
ET PPV	0.5634	0.8154	0.7583	0.8127	0.7991
WT Sensitivity	0.7986	0.8595	0.8296	0.8553	**0.8818**
TC Sensitivity	0.8082	0.9069	0.8739	0.9102	**0.9145**
ET Sensitivity	0.6329	0.8203	0.7698	0.8099	**0.8453**
WT Hausdorff	3.3011	2.5787	2.7640	2.6126	2.5804
TC Hausdorff	2.2013	1.6516	2.0579	1.7696	**1.5660**
ET Hausdorff	3.6041	2.7487	3.0706	2.8314	**2.7331**

3.4. Baseline Methods

We compare our model with the following baseline methods:

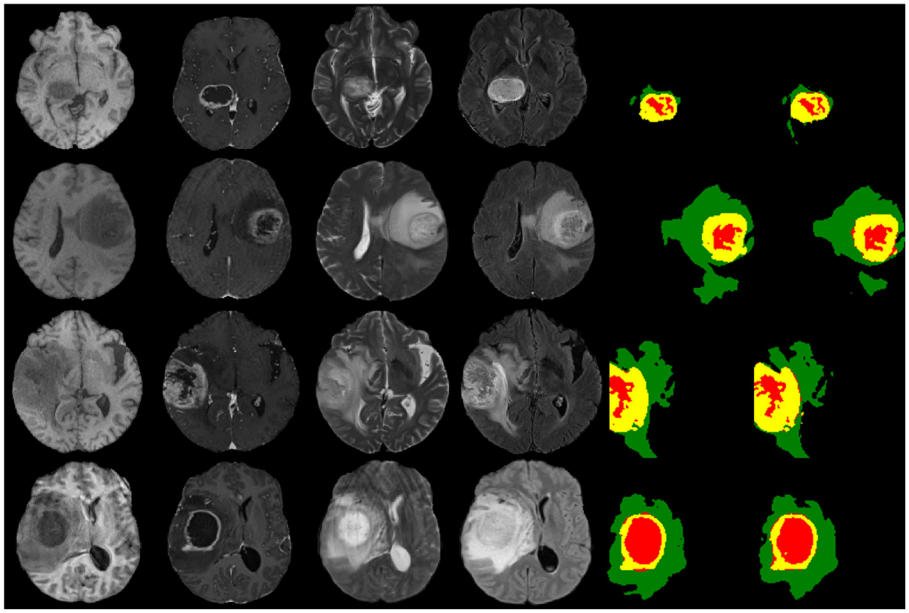


FIGURE 9 | Segmentation results for four cases. C_1 are the images represent one axial MRI slice obtained in T1-weighted, C_2 are the T1c-weighted, C_3 are the T2-weighted, C_4 are the FLAIR-weighted, C_5 are the the standard tumor segmentation results manually labeled by experts, and the C_6 show the segmentation results obtained in this section.

FCN32s (Shelhamer et al., 2016): It can accept input image of arbitrary size, and use a deconvolution layer to upsample the feature map of the last convolution layer. It recovers images to the same size as the input image, which can produce a prediction for each pixel while preserving the spatial information in the original input image.

UNet (Ronneberger et al., 2015): The main idea is to connect a network with a similar structure to the previous layer behind the downsampling stage, and then recover the resolution of the image output by applying upsampling. Finally combine the upsampled output with the downsampled output which has high-resolution features. By these steps the edge features of the image can be better extracted.

Attention-UNet (Oktay et al., 2018): The attention mechanism is introduced to learn the importance of each element with respect to the target. It limits the activation to the region with segmentation and reduces the activation value of the background to optimize the result.

ResNet-UNet: UNet with residual block helps to solve the gradient vanishing and gradient exploding problems and train deeper networks while ensuring good performance.

3.5. Performance Comparison

Table 3 and **Figure 8** summarize the baseline results compared to our proposed methodology. It is apparent from **Table 3** that our proposed framework outperforms the other four approaches across 9 metrics and achieves the highest Dice and Sensitivity, and the results are shown in bold values.

Figure 9 shows the segmentation results for the four cases obtained from the segmentation network of our proposed model. In these graphs, each row represents a real clinical case. From left to right, the images represent one axial MRI slice obtained in T1-weighted, T1c-weighted, T2-weighted, and FLAIR-weighted, the standard tumor segmentation results manually labeled by experts, and the last column shows the segmentation results obtained in this section. The tumor categories are highlighted with different colors: enhancing tumor regions (yellow), peritumor edema regions (green), and necrotic and non-enhancing tumors (red). It can be seen that the tumor size, shape, location, and category differ in the four brain MRIs. As can be seen from the comparison graph, the segmentation network of the Dense-ResUNet model is close to the standard segmentation results of cancer tumors manually labeled by experts, which proves the effectiveness of the Dense-ResUNet model.

4. CONCLUSION

In this article, we design a model based on 2D UNet brain tumor segmentation model and named it as Dense-ResUNet. By combining the features of multi-modal brain tumor MRI images, the attributes of each phase in MRI images were improved. The

Dense-ResUNet makes full use of the nested dense convolutional blocks to fill in the gaps of traditional UNet, which capture different levels of features. Then it uses a residual unit to extract pixel information from the image and skip the link to solve the traditional deep CNN network problem. Finally the multi-scale feature maps are fused to obtain the segmentation results. Experiment results show that our proposed method is better than other comparative approaches. The result can prove the effectiveness of our framework and show that the Dense-ResUNet can help extract the fusion of multi-modal of brain tumor images through the image convolutional network, which has clinical research and application value.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://aistudio.baidu.com/aistudio/datasetdetail/64660.

AUTHOR CONTRIBUTIONS

SC wrote the main manuscript and conducted the experiments. SZ and QL wrote, reviewed, edited, and supervised the manuscript. All authors read and approved the submitted manuscript.

REFERENCES

Abdelmaksoud, E., Elmogy, M., and Al-Awadi, R. (2015). Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Inf. J.* 58, 71–81. doi: 10.1016/J.EIJ.2015.01.003

Aslam, A., Khan, E., and Beg, M. (2015). Improved edge detection algorithm for brain tumor segmentation. *Procedia Compu. Sci.* 58, 430–437. doi: 10.1016/j.procs.2015.08.057

Brosch, T., Yoo, Y., Tang, L.Y.W., Li, D.K.B., Traboulsee, A., and Tam, R. (2015).“Deep convolutional encoder networks for multiple sclerosis lesion segmentation,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, Vol. 9351*, eds N. Navab, J. Hornegger, W. Wells, and A. Frangi (Cham: Springer). doi: 10.1007/978-3-319-24574-4_1

Hartigan, J. A., and Wong, M. A. (2013). A k-means clustering algorithm. *Appl. Stat.* 28, 100–108. doi: 10.2307/2346830

Hou, L., Samaras, D., Kurc, T., Gao, Y., Davis, J., and Saltz, J. (2016). “Patch-based convolutional neural network for whole slide tissue image classification,” in *Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016 (Las Vegas, NV).

Kaus, M., Warfield, S., Nabavi, A., Black, P., Jolesz, F., and Kikinis, R. (2001). Automated segmentation of mr images of brain tumors. *Radiology* 218, 586–591. doi: 10.1148/radiology.218.2.r01fe44586

Kittler, J., and Illingworth, J. (1986). Minimum error threshold. *Pattern Recognit.* 19, 41–47.

Legg, P., Rosin, P., Marshall, D., and Morgan, J. (2015). Feature neighbourhood mutual information for multi-modal image registration: an application to eye fundus imaging. *Pattern Recognit.* 48, 1937–1946. doi: 10.1016/j.patcog.2014.12.014

Mahmood, Q., and Basit, A. (2016). “Automatic ischemic stroke lesion segmentation in multi-spectral mri images using random forests classifier,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, eds A. Crimi, B. Menze, O. Maier, M. Reyes, and H. Handels (Cham: Springer). doi: 10.1007/978-3-319-30858-6_23

Milletari, F., Navab, N., and Ahmadi, S. (2016). “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 Fourth International Conference on 3D Vision (3DV)*, 565–571. doi: 10.1109/3DV.2016.79

Moeskops, P., Viergever, M., Mendrik, A., Vries, L., Benders, M., and Isgum, I. (2016). Automatic segmentation of mr brain images with a convolutional neural network. *IEEE Trans. Med. Imag.* 35, 1252–1261. doi: 10.1109/TMI.2016.2548501

Nandpuru, H., Salankar, S., and Bora, V. (2014). “MRI brain cancer classification using support vector machine,” in *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science* (IEEE), 1–6. doi: 10.1109/SCEECS.2014.6804439

Oktay, O., Schlemper, J., Folgoc, L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). Attention u-net: Learning where to look for the pancreas. arXiv [Preprint]. arXiv: 1804.03999. Available online at: https://arxiv.org/pdf/1804.03999.pdf (accessed May 20, 2018).

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, Vol. 9351*, eds N. Navab, J. Hornegger, W. Wells, and A. Frangi. doi: 10.1007/978-3-319-24574-4_28

Rtsch, G., Onoda, T., and Mller, K.-R. (2001). Soft margins for adaboost. *Mach. Learn.* 42, 287–320. doi: 10.1023/A:1007618119488

Saad, N. M., Bakar, S. A. R. A., Muda, S., and Mokji, M. (2010). “Automated segmentation of brain lesion based on diffusion-weighted mri using a split and merge approach,” *Proceedings of 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2010* (Kuala Lumpur).

Shelhamer, E., Long, J., and Darrell, T. (2016). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1. doi: 10.1109/TPAMI.2016.2572683

Shi, T., and Horvath, S. (2005). Unsupervised learning with random forest predictors. *J. Comput. Graph. Stat.* 15, 118–138. doi: 10.1198/106186006X94072

Sobel, I. (1970). *Camera models and machine perception* (Ph.D. dissertation). Stanford University, United States.

Tustison, N., Shrinidhi, K., Wintermark, M., Durst, C., Kandel, B., Gee, J., et al. (2014). Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with antsr. *Neuroinformatics* 13, 209–225. doi: 10.1007/s12021-014-9245-2

Vaughan, J., Snyder, C., DelaBarre, L., Tian, J., Akgun, C., Ugurbil, K., et al. (2006). “Current and future trends in magnetic resonance imaging (MRI),” in *2006 IEEE MTT-S International Microwave Symposium Digest*, 2006, pp. 211–212, doi: 10.1109/MWSYM.2006.249451

Zhang, H., Berg, A., Maire, M., and Malik, J. (2006). “SVM-KNN: discriminative nearest neighbor classification for visual category recognition,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE), 2126–2136. doi: 10.1109/CVPR.2006.301

Zhou, Z., Siddiquee, Mmr, Tajbakhsh, N., and Liang, J. (2018). “UNet++: a nested U-net architecture for medical image segmentation,” in *4th Deep Learning in Medical Image Analysis (DLMIA) Workshop* (Granada).

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, Zhao 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 in Neuroscience

ORIGINAL RESEARCH
published: 22 March 2022
doi: 10.3389/fnins.2022.836412

Cross-Attention and Deep Supervision UNet for Lesion Segmentation of Chronic Stroke

Manjin Sheng[1], Wenjie Xu[1], Jane Yang[2] and Zhongjie Chen[3]**

[1] School of Informatics, Xiamen University, Xiamen, China, [2] Department of Cognitive Science, University of California, San Diego, San Diego, CA, United States, [3] Department of Neurology, Zhongshan Hospital, Xiamen University, Xiamen, China

Stroke is an acute cerebrovascular disease with high incidence, high mortality, and high disability rate. Determining the location and volume of the disease in MR images promotes accurate stroke diagnosis and surgical planning. Therefore, the automatic recognition and segmentation of stroke lesions has important clinical significance for large-scale stroke imaging analysis. There are some problems in the segmentation of stroke lesions, such as imbalance of the front and back scenes, uncertainty of position, and unclear boundary. To meet this challenge, this paper proposes a cross-attention and deep supervision UNet (CADS-UNet) to segment chronic stroke lesions from T1-weighted MR images. Specifically, we propose a cross-spatial attention module, which is different from the usual self-attention module. The location information interactively selects encode features and decode features to enrich the lost spatial focus. At the same time, the channel attention mechanism is used to screen the channel characteristics. Finally, combined with deep supervision and mixed loss, the model is supervised more accurately. We compared and verified the model on the authoritative open dataset "Anatomical Tracings of Lesions After Stroke" (Atlas), which fully proved the effectiveness of our model.

Keywords: lesion segmentation, chronic stroke, deep learning, MRI, ATLAS

OPEN ACCESS

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Zhijin Wang,
Jimei University, China
Ziquan Zhu,
University of Leicester,
United Kingdom

****Correspondence:***
Jane Yang
j7yang@ucsd.edu
Zhongjie Chen
fmuzhongjie@aliyun.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *17 December 2021*
Accepted: *26 January 2022*
Published: *22 March 2022*

Citation:
Sheng M, Xu W, Yang J and Chen Z (2022) Cross-Attention and Deep Supervision UNet for Lesion Segmentation of Chronic Stroke. Front. Neurosci. 16:836412. doi: 10.3389/fnins.2022.836412

INTRODUCTION

According to the definition of the World Health Organization, stroke is a sudden condition caused by blood vessels in the brain. The brain is the overall control center of various activities of the human body so stroke can lead to disability or death. The incidence rate of stroke has been increasing in recent years. Therefore, the study of cerebral apoplexy has become a popular direction of clinical medicine. In the past, stroke has become a leading cause of death and the second leading cause of disability after ischemic heart disease (Feigin, 2019). With the development of society, due to the changes in the living environment and corresponding lifestyle changes, the number of stroke patients is increasing daily. Hypertension, poor diet, poor physical activity habits, diabetes, obesity, dyslipidemia, as well as psychosocial stress, economic status, air pollution, and other factors, are risk factors for stroke (Pandian et al., 2018). These more common factors have made stroke a common disease with an increasing number of patients. The study of Feigin, VL (Feigin, 2019) showed that the morbidity and mortality of stroke victims had increased significantly since 2010, and the number of strokes in 2016 was twice that in the 1990s.

Stroke is mainly divided into ischemic stroke and hemorrhagic stroke, which are caused by vascular obstruction and vascular rupture, respectively. Once cerebral apoplexy occurs, it may

cause permanent damage to the brain and lead to a variety of complications, which can easily lead to disability or even death of patients. Patients who recover are also prone to relapse, therefore, stroke not only affects the normal life of patients, but even endangers the life and health of patients. In addition, the occurrence of stroke will also bring a certain burden to the family and society. Therefore, timely and effective detection and identification of the location of the disease is the basis of minimizing injury. With the development of medical imaging technology, the clinical treatment and intervention of stroke patients has become convenient, accurate, and efficient through the assistance of medical imaging. The identification and analysis of disease location of cerebral apoplexy by medical imaging are now the main diagnostic method in clinical practice. There are three imaging methods for damaged brain tissue of stroke patients: CT angiography (CTA), magnetic resonance imaging (MRI), and digital subtraction angiography (DSA). CT imaging performed well in the diagnosis of tumor-induced stroke patients, but the imaging results of brain regions other than the brain and early lesions were weak. Meanwhile, angiography requires the injection of a contrast agent into the patient, which has implications for the patient's body. Non-invasive brain MRI results, by contrast, are not only for ischemic stroke imaging but is more sensitive, and can accurately analyze the lesions. In addition, for hemorrhagic stroke in the different periods of the hemorrhage of magnetic resonance imaging results, through multiple check sequences, it can realize direct imaging of multidimensional no dead angle, which provides favorable conditions for the doctor's clinical diagnosis and treatment. Therefore, this paper focuses on T1-Weighted Magnetic Resonance Images of cerebral apoplexy patients to study the location segmentation of lesions in the images.

At present, the number of studies on the segmentation of acute stroke is much larger than that on the segmentation of chronic stroke, and the study of chronic stroke is of great significance for early intervention treatment and prevention of serious consequences. According to whether manual participation is needed in the algorithm, the segmentation methods of stroke lesions can be divided into semi-automatic and automatic categories at present. Clustering is the most widely used semi-automatic segmentation method. Haan et al. (2015) segmented through the iterative region growing based on local intensity maximum. However, this method not only requires a certain amount of manpower because of the need for artificial selection of clusters, but also its effect is greatly influenced by human experience. In automatic segmentation, machine learning is also a mainstream method. For example, Seghier et al. (2008) used fuzzy means clustering to identify the location of brain lesions. However, the application scope of this method is limited, and it is only applicable to single types of images. The sensitivity of this method is limited by parameters and other factors. Griffis et al. (2016) classified voxels by Gaussian Naive Bayes, but this method is prone to miss detection for small lesion areas. Pustina et al. (2016) used the random forest algorithm and combined it with adjacent voxels for analysis, but this method had the defect of high computational cost. Meanwhile, Ito et al. (2019) conducted an experimental comparison of the above three methods, and the results showed that these methods had limitations in the segmentation of lesions in the cerebellum and brain stem.

In addition, the experiment of Maier et al. (2015) also showed that machine learning had a good performance in the segmentation task of stroke lesions. In machine learning methods, in addition to model-based Gaussian Naive Bayes, random forest, and k-neighborhood algorithm, there are also deep learning methods taking convolutional neural network (CNN) as an example. Deep learning methods have many mature and robust methods in the field of medical image classification (Lyu et al., 2021; Zhu et al., 2021a,b). At the same time, there has been much development in the field of medical image segmentation. Among them, CNN and full convolutional neural network (FCN) show obvious advantages in the segmentation of lesions in brain images (Karthik et al., 2020). Lucas et al. (2017) applied the classical fully connected neural network to the segmentation of stroke lesions. In addition, U-Net architecture also shows a good effect in image segmentation, so there are many applications of FCN, U-Net, and their variants in the segmentation task of stroke lesions. Liu et al. (2018) proposed a fully convolutional network with a residual structure (Res-FCN). Hui et al. (2020) combined the attention mechanism with U-Net and improved it to increase the segmentation accuracy by suppressing the irrelevant regions of lesions through the attention generated by multi-scale features. Dolz et al., 2019) obtained more modal information through individual processing of the input image and extended the convolution block of InceptionNet to obtain more context information to improve performance. However, U-Net architecture is limited by the size of convolutional features, and the simple stacking of features during decoding limits the diversity of information. Therefore, Qi et al. (2019) proposed an X-Net based on deep separable convolution, which designed a non-local operation, namely feature similarity module (FSM), to capture remote dependencies. Using deep convolution can reduce the scale of the network, and the developed FSM can provide more effective and dense context information extraction, which is conducive to better segmentation.

Unfortunately, these methods often ignore the size and location information of lesions, do not make full use of the extracted feature information, and lose the early useful features. Therefore, we propose an end-to-end model named cross-attention and deep supervision UNet (CADS-UNet), which refers to the classical UNet structure, designs the cross-space attention module (CSAM) and the channel attention module (CAM), filters the features of each layer, and extracts more valuable features. In summary, the contributions of this paper are summarized as follows.

(1) We propose CADS-UNet to effectively segment the lesion region of stroke. Through the cross-spatial attention module and CAM, important information is enhanced and the effectiveness of features is improved.

(2) We use a mixed loss function and deep supervision to improve the efficiency of network training and provide strong normalization through supervising the middle layer.

(3) We conducted comparative experiments and ablation experiments on the Atlas dataset and compared with advanced methods to show the advantages.

MATERIALS AND METHODS

Figure 1 shows the overall architecture of our proposed CADS-UNet. On the general encoder-decoder architecture, we add the cross spatial attention module and CAM to effectively select features, fully mix features from different levels, and guide segmentation. We utilize ResNet34 (He et al., 2016) as the encoder, which contains a total of five blocks. Each decoder block consists of two Conv-BN-ReLU combinations. At the same time, the deep supervision algorithm is combined to improve the directness and transparency of the hidden layer learning process. In this section, each module of our network will be analyzed in detail. In section "Cross-Spatial Attention Module," we introduce the cross spatial attention module (CSAM), the CAM in section "Channel Attention Module," and finally, in section "Deep Supervision and Mixing Loss Function," we introduce the deep supervision and mixing loss function.

Cross-Spatial Attention Module

Convolutional neural network extracts the features of the target by layer abstraction and one of the important concepts is a receptive field. The receptive field of a low-level network is relatively small and the representational ability of geometric detail information is strong. It tends to extract low-level features such as the edgy texture of an image. Although the resolution of low-level features is high, the representational ability of semantic information is weak. With the increase of neural network layers, the receptive field increases, and the representational ability of semantic information is strong, but the resolution of the feature map is low and the representational ability of geometric information is weak. The classical U-Net performs up-sampling four times and uses skip connection in the same stage to combine the encoded feature map with the decoded feature map so that the feature map integrates more low-level features and the features of different scales. Due to the spatial uncertainty of stroke lesions, to obtain more effective spatial features, we propose a CSAM to replace the skip connection in U-Net. This module extracts a wide range of location-sensitive information at encoding and decoding layers, respectively, and cross-encodes them into feature map.

CSAM is a soft spatial attention strategy (Tao et al., 2018) that selects fine-grained important pixels which are pixel level. The purpose is to combine features with spatial location strategy, so that the obtained features pay more attention to uncertain and complex regions, to realize hierarchical feature

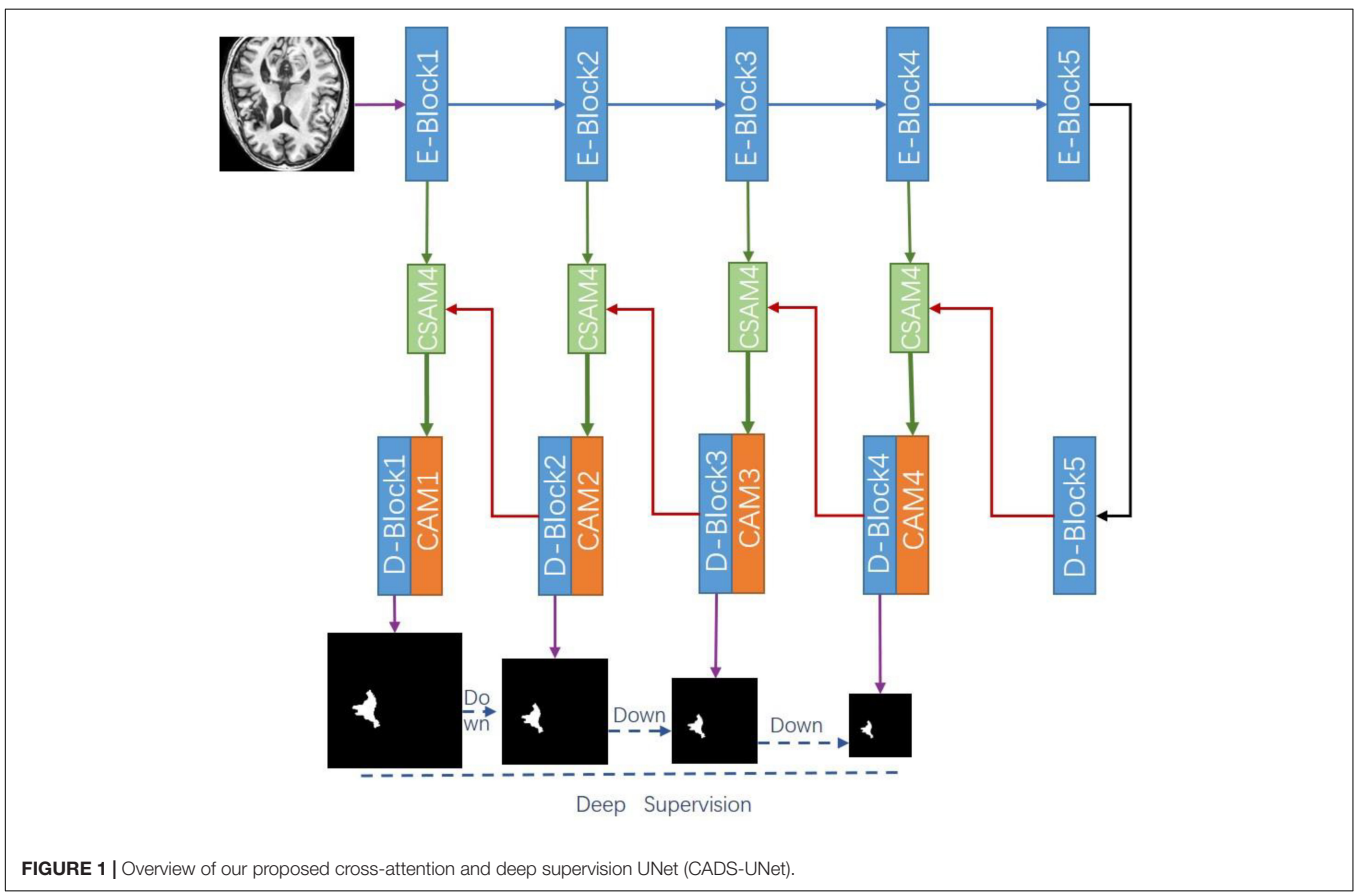


FIGURE 1 | Overview of our proposed cross-attention and deep supervision UNet (CADS-UNet).

complimentary and prediction refinement. Different from the common self-attention module (Park et al., 2018), we select features by cross-weighting the spatial features of the encoding layer and decoding layer. The features of the coding layer include unique edge space features, while the features of the decoding layer include high-level abstract space features. Both sides focus on the features of different levels and lose the spatial focus of the other layer. Directly connecting encoding and decoding features cannot efficiently select spatial information. Therefore, we use the soft attention mechanism to cross weight the features, so that the features are concentrated in different key parts of space to prevent the loss of effective information and strengthen important information.

As shown in **Figure 2**, the attention diagram of each CSAM module is generated by the decoder and encoder of the corresponding layer. The input feature maps of CSAM are the encoding feature $F_e \in \mathbb{R}^{C\times H\times W}$ and the decoding feature $F_d \in \mathbb{R}^{C\times H\times W}$, in which C is the number of channels and H, W are the height and width of the feature map, respectively. To compute the spatial attention, first applying max pooling and avg pooling operations along the channel axis to obtain two 2D maps, then concatenate them according and send the concatenated feature descriptor to a standard convolution layer. After activating function, the spatial attention map is obtained, and its formula is expressed as:

$$\boldsymbol{M}\ (\boldsymbol{F})\ = \sigma\left(\boldsymbol{f}^{\mathbf{3}\times\ \mathbf{3}}\left(\left[\boldsymbol{AvgPool}\left(\boldsymbol{F}\right);\boldsymbol{MaxPool}\left(\boldsymbol{F}\right)\right]\right)\right)$$

Where σ represents the sigmoid function and $\boldsymbol{f}^{\mathbf{3}\times\mathbf{3}}$ denotes a convolution operation with the filter size of $\mathbf{3}\times\ \mathbf{3}$.

The spatial attention maps of F_e and F_d are extracted, respectively, and the original features are weighted by cross-multiplication to obtain a hybrid spatial selection feature map. Finally, the original features are summed as residual blocks to avoid overfitting. The specific formula is as follows:

$$\boldsymbol{S_S}\left(\boldsymbol{F_e},\boldsymbol{F_d}\right)\ =\left[\left(\boldsymbol{M}\ \boldsymbol{F_e}\otimes\boldsymbol{F_d}+\boldsymbol{F_d}\right);\left(\boldsymbol{M}\ \boldsymbol{F_d}\otimes\boldsymbol{F_e}+\boldsymbol{F_e}\right)\right]$$

Where $\otimes$ denotes element-wise multiplication, it can be inferred from the formula that the final result map is the sum of relational features and original features. Therefore, S_s has a wide range of contextual perspectives, effectively aggregates spatial information through the cross-spatial attention module, provides local context enhancement of different receptive fields for each position feature column of each decoding layer, and skillfully uses the prediction confidence of the decoding layer as a guide to force the current layer to focus on difficult.

Channel Attention Module

Figure 3 shows the CAM. In a sense, the channel of a neural network is the feature, and the number of channels is the feature number. In addition to spatial location information, channel information is also an important information source for selecting effective feature maps. The mixed attention mechanism is better than the single attention mechanism (Woo et al., 2018). So, we use the CAM serially after the CSAM to select the combined multi-feature channels. We hope to improve the representation ability of the network modeling the dependence of each channel

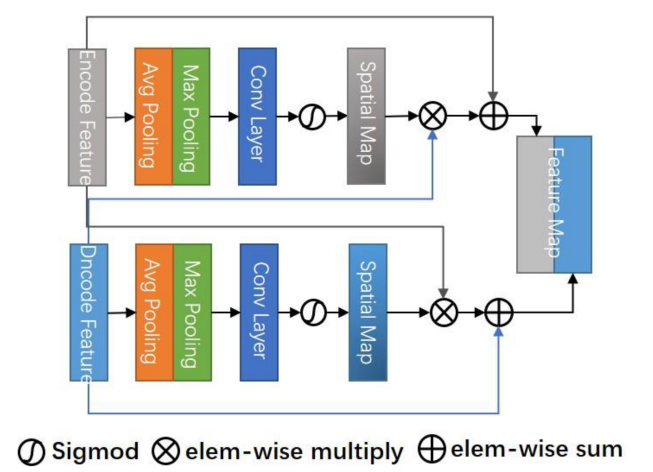


FIGURE 2 | Cross space attention module (CSAM).

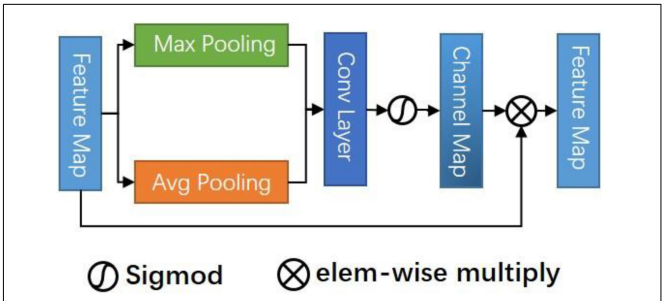


FIGURE 3 | Channel attention module (CAM).

and adjusting the features channel by channel, in this way, the network can learn to selectively strengthen the features containing useful information and suppress useless channel features globally.

Specifically, CAM takes the connected mixed features S_s F_e, F_d as inputs to suppress or enhance the channel strength by modeling the importance of each channel. Firstly, max pooling and avg pooling are used to summarize the spatial information about the feature graph. We obtain two descriptors and then forward them to a shared multi-layer perceptron (MLP) that contains two layers of FC, respectively. Finally, the pixels are added and fused, and the channel attention map is obtained through an activation function. The formula is:

$$\boldsymbol{M}\left(\boldsymbol{S_s}\right)\ =\sigma\left(\boldsymbol{MLP}\left(\boldsymbol{AvgPool}\left(\boldsymbol{S_s}\right)\right)+\boldsymbol{MLP}\left(\boldsymbol{MaxPool}\left(\boldsymbol{S_s}\right)\right)\right)$$

By multiplying the original feature mapping with the attention value, we can identify some informative contextual features and suppress those features that are not conducive to improving the discrimination.

$$\boldsymbol{S_c}\left(\boldsymbol{S_s}\right)\ =\boldsymbol{M}\ \boldsymbol{S_s}\otimes\boldsymbol{S_s}$$

Deep Supervision and Mixing Loss Function

We add cross-space and channel dual attention modules in each decoding layer. In addition, we use deep supervision (Lee et al., 2014) to introduce intermediate loss supervision functions for each hidden layer to improve the directness and transparency of the hidden layer learning process. These intermediate loss functions can be regarded as additional (soft) constraints on the learning process. At the same time, it can solve the problems such as the disappearance of training gradient and slow convergence speed of deep neural networks.

In the training process, deep supervision may lead to lazy convergence of the deep network. Therefore, we propose a new loss function, which adopts different loss functions according to the output of each layer and combines them.

Binary Cross Entropy Loss

For the segmentation problem of front and back scene pixel-level classification, the most widely used objective function is the binary cross-entropy (BCE) loss, which is defined as follows:

$$L_{BCE} = \frac{1}{N}\sum_{i=1}^{N} y_i log\widehat{y}_i + (1 - y_i)\, log\,(1 - \widehat{y}_i)$$

Where N is the total number of pixels, y_i is the real label of pixel i, and $\hat{y}_i$ represents the prediction probability that pixel i is classified as "1." This loss function ignores class imbalance, so the importance of all pixels is the same. In this case, the larger background has more influence on the training process.

Dice Coefficient Loss

In 3D medical data, including our MRI stroke data, the number of lesion slices only accounts for a small part of the total number of 3D slices, and the size of lesions is uncertain. In the training process, the number of negative pixels and slices (excluding lesions) is much more than the number of positive pixels, and a large number of negative regions may dominate the loss function. Therefore, using BCE loss alone cannot meet our task. Dice coefficient loss (DL) (Milletari et al., 2016) does not take into account the real negative pixels, alleviates the imbalance between background and foreground pixels, and can achieve better results in unbalanced segmentation. Sim ilar to Dice score, dice loss is defined as follows:

$$L_{Dice} = 1 - \frac{2\sum_i y_i\widehat{y}_i + \epsilon}{\sum_i \widehat{y}_i + \sum_i y_i + \epsilon}$$

Where $\epsilon \in [0, 1]$ is a tunable parameter to prevent a divide-by-zero error and make negative samples also have gradient propagation.

Mixed Loss

Based on the above two losses, we propose a mixed loss (ML) to improve the convergence. Combined with the deep supervision strategy, we use the dice coefficient loss function for the coarser step 3 and 4 outputs, and $L_{BCE} + L_{Dice}$ loss function for the finer step 1 and 2 outputs. For different strides, different loss functions are used, because the coarse output only focuses on the global structure and ignores the local details, while the fine output achieves pixel-level accuracy by relying on local clues, and we use α, β to add weights for different layers. Finally, the mixed loss is as follows:

$$L = \left(L^1_{Dice} + L^1_{BCE}\right) + \alpha\left(L^2_{Dice} + L^2_{BCE}\right) + \beta\left(L^3_{Dice} + L^4_{Dice}\right)$$

EXPERIMENTS AND RESULTS

Datasets

To evaluate the performance of the proposed method in the segmentation of stroke lesions, we used a subset of the open data set "Anatomical Tracings of Lesions After Stroke" (Atlas) (Liew et al., 2018). Atlas dataset consists of 304 MRI images from 11 cohorts around the world. A total of 11 experts who had received standardized training in lesion recognition and segmentation manually drew lesion masks for the data set. Due to the difference in technical difficulty and scanner image quality, 229 brain data were converted into standard MNI space. Therefore, experiments are only carried out on these 229-subset data.

All T1 weighted MRI data are collected on a 3T MRI scanner with a resolution of $1mm^3$(isotropic). The mean lesion volume of all cohorts is $2.128 \pm 3.898 \times 10^4$ mm^3. The minimum lesion size is 10 mm^3 and the maximum lesion size is 2.838×10^5 mm^3. On average, a single patient is more likely to have only one lesion (58%). In the left hemisphere, right hemisphere, and other parts (such as the brain stem), the probability of identifying lesions is roughly the same (48.4, 43.8, and 7.7%, respectively). Overall, the number of subcortical lesions (70.7%) is higher than that of cortical lesions (21.5%).

The experimental data consists of 229 T1 weighted standardized 3D MR images, in which each 3D MR image consists of 189 2D slices, and the size of each slice is 233 × 197. The 2D segmentation method was used in this experiment, so the slices are standardized. The data set contains 43281 slices. We randomly selected 137 cases for training, 36 cases for verification, and 56 cases for testing.

Implementation Details

In data preprocessing, each case was cut into a cross-sectional image. Each image was cropped using diagonal coordinates (10,40) and (190,220). The cropped region removed irrelevant information and expanded the proportion of stroke lesions in the whole image. Next, the size of the cropped image was adjusted to 192 × 192 using bilinear interpolation, and saved each slice in.npy format for later use. The training data were shuffled randomly and all slice masks were cut to the same size.

In this experiment, Adam (Kingma and Ba, 2014) optimizer was used to adjust the initial learning rate and the initial learning rate was set to $init_lr = 1e^{-3}$. We adjusted the initial learning rate by using multi learning rate strategy, which was $lr = init_lr \times \left(1 - \frac{epoch}{nEpoch}\right)^{power}$, where $power = 0.9$, $nEpoch = 100$. Due to the memory limitation of GPU, the batch size was set to 24.

Our model was implemented using the PyTorch (Paszke et al., 2019) framework.

We selected a series of evaluation indicators to quantify the performance of the proposed model. We calculated evaluation scores for each patient in the test set on 3D MRI (189 2D slices) and reported the mean Dice Similarity Coefficient (DSC), Precision, and Recall. In addition, we also tested the global DSC of all data. DSC is the main indicator to evaluate the difference between the model prediction and the ground truth, which is used to quantify the overlap ratio between the two masks. Precision, also known as the positive predictive value, is the ratio of true positive to the overall predictive result. The recall is an integrity measure defined as the ratio of true positive to the entire basic fact. They are defined as:

$$DSC = \frac{2TP}{2TP + FP + FN}$$

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

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

We threshold all predicted results. We set it to 0 when the probability of a pixel being predicted to be foreground is less than 0.5, and 1 otherwise. It is positive when the pixel predicted value is the same as the true value, otherwise it is negative. Where true positive (TP) indicates that the model correctly predicted pixel. False-positive (FP) indicates the pixel that the model misclassifies to be positive. False-negative (FN) indicates that the positive pixel is mistakenly classified to be negative by the model.

Comparison of Different Methods

To evaluate the performance of our algorithm, we selected four models for experimental comparison (**Table 1**), namely FCN-8s (Long et al., 2015), U-Net (Ronneberger et al., 2015), ResUNet (Zhang et al., 2018), and Attention-UNet (Oktay et al., 2018). As the baseline of the segmentation network, the U-Net network has a simple and clear structure and has a good effect. Subsequent networks, including our network, are improved based on U-Net. ResUNet combines the idea of ResNet and adds a residual module to each block. Attention-UNet monitors the features of the upper level through the features of the lower level to realize the attention mechanism. All network parameters are configured according to section "Implementation Details." These networks are also implemented on a unified platform. The training set and verification sets are strictly consistent (including data preprocessing). During training, the batch size is set to 24 and the loss is set to $L_{Dice} + L_{BCE}$.

TABLE 1 | Comparison with state-of-the-art methods on the ATLAS dataset.

Method	DSC	DSC (global)	Recall	Precision
FCN-8s	0.4274	0.6720	0.4531	0.4883
U-Net	0.4944	0.6933	0.5030	0.6419
ResUNet	0.5081	0.7373	0.5081	0.5921
Attention-UNet	0.5162	0.7383	0.5321	**0.6518**
CADS-UNet (ours)	**0.5564**	**0.7451**	**0.5817**	0.6368

Bold value shows the best performance.

By analyzing the results in **Table 1**, we can see that the effect of classical FCN-8s is the worst, only 0.4274 (DSC). Taking the result of U-Net 0.4944 (DSC) as the dividing line, the results of ResUNet and Attention-UNet are improved. The best result of Attention-UNet is 0.5162 (DSC), which also proves the usefulness of the attention mechanism. Our model obtained a DSC of 0.5564, which scored the highest among all models, and several other indicators [DSC (global), Recall, Precision] were 0.7451, 0.5817, and 0.6368, respectively. Therefore, from the experimental point of view, our model is optimal.

Figure 4 shows some of our cases. We arrange the images from small to large according to the size of the lesion. Where the second column is the ground truth and our resulting graph is in the third column. In the first row, we can see that for very small lesion areas, U-Net and ResUNet cannot segment the lesion, the predicted mask is empty, and the segmentation result of FCN-8s is not good. In contrast, our model and Attention-UNet can accurately segment the lesion area. Therefore, compared with other models, our method has advantages in identifying and segmenting small lesion samples. In addition, take the result in line 5 as an example, U-Net, ResUNet, and Attention-UNet mistakenly recognized the tissue with the low-intensity signal as the lesion area of stroke and the segmentation is not accurate enough. Our model does not segment the non-lesion area and the segmentation shape is most similar to the ground truth. Other results also show that our model obtains smoother edge results and a more accurate segmentation rate. The experimental results show that the proposed CADS-UNet has strong robustness and effectiveness.

Comparison With Other Methods and Ablation Experiments

Table 2 compares our model (CADS-UNet) with other existing methods on the same data set. All comparison results were obtained directly from the paper. We did not reproduce the experiment because we believe in the best results reported by other authors. We selected five methods based on deep learning (X-Net, 2D MI-UNet, 2.5D CNN, D-UNet, and 3D-UNet). The results of X-Net and D-UNet were obtained from Qi et al. (2019) and Zhou et al. (2019). The results of 2.5D CNN were obtained from Xue et al. (2020) and the results of 3D-UNet and 2D MI-UNet were reported by Zhang et al. (2020). As can be seen from the Table, compared with other methods, our model also has certain advantages.

In order to verify the necessity and effectiveness of each module (CSAM, CAM, DL), we conducted ablation experiments between different modules. Specifically, the baseline model refers to the encode-decode network, the feature connection adopts

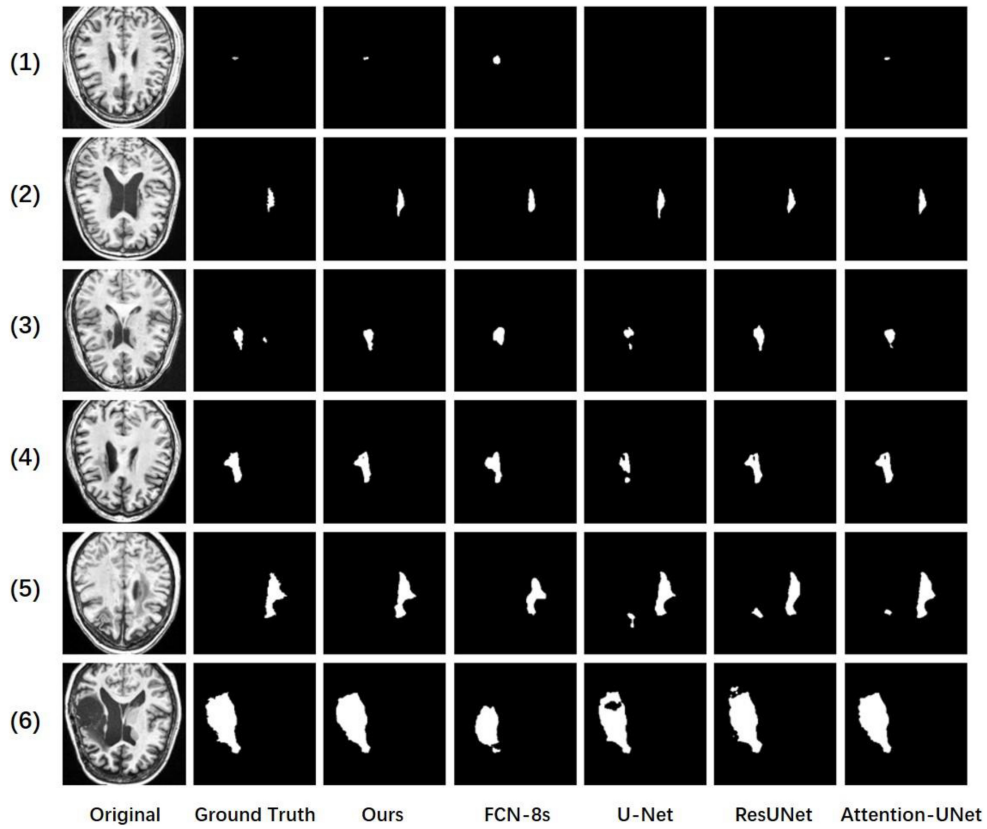


FIGURE 4 | Comparisons of our method, baseline, FCN-8s, U-Net, ResUNet, and attention-UNet.

connection instead of direct addition to double the number of channels. Before adding the deep supervision module, our loss adopts $L_{Dice} + L_{BCE}$. We gradually add CSAM, CAM and DL modules to the base model, which are expressed in the table as baseline, BL + CSAM, BL + CSAM + CAM, and ours (BL + CSAM + CAM + DL). From the results, we can see that with the introduction of the module, the effect of our algorithm has been gradually improved, boosting dice by 2.37, 0.46, and 1.57%, respectively. The addition of the CSAM module greatly improves the accuracy, which also proves that the model does capture the corresponding spatial location information, and the joint use of CAM and CSAM also improves the model. Finally, a specific loss function constrains the output of each layer, making the focus of the attention mechanism more accurate, so that the whole model reaches the optimum. The experimental results show that our added module is indeed effective.

TABLE 2 | Comparison with other state-of-the-art methods and ablation studies on the ATLAS dataset.

Method	DSC	Recall	Precision	Train/Test
X-Net	0.4867	0.4752	0.6000	All/Fivefold cross-validation
2D MI-UNet	0.4945	0.5237	0.5669	All/Fivefold cross-validation
3D UNet	0.5296	0.5497	0.6090	All/Fivefold cross-validation
D-UNet	0.5349	0.5243	0.6331	183/46
2.5D CNN	0.54	–	–	99/Fivefold cross-validation
CADS-UNet (ours)	**0.5564**	**0.5817**	0.6368	137/56
BaseLine (BL)	0.5124	0.5291	0.6111	137/56
BL + CSAM	0.5361	0.5455	**0.6451**	137/56
BL + CSAM + CAM	0.5407	0.5654	0.6218	137/56

Bold value shows the best performance.

CONCLUSION

We propose a new method to realize the automatic segmentation of stroke lesions. The network strengthens the feature space information through the cross-spatial attention module so that the model pays more attention to uncertain and complex regions. At the same time, the channel information is filtered through the CAM, the effective channel information is strengthened, and more accurate feature information is captured. Finally, the deep supervision strategy combined with the mixed loss function solves the problems of the disappearance of the training gradient and the slow convergence speed of the deep neural network. The model is evaluated on the famous open data set ATLAS, which proves the superiority of our model. The proposed model has a clear structure and robustness. In the future, we plan to

validate our method on more clinical datasets and improve the model in combination with 3D convolutions to further improve the performance of the model. Due to the continuity of the lesion block, 3D convolution can add spatial information to the model, which can theoretically further improve our segmentation accuracy, but 3D convolution has higher requirements for computing performance. The 2.5D network that combines 2D and 3D is also our future research direction.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in the ATLAS dataset: http://fcon_1000.projects.nitrc.org/indi/retro/atlas_download.html.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. 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

MS wrote the main manuscript and conducted the experiments. WX and JY participated in the writing of the manuscript and modified the English grammar of the manuscript. JY and ZC contributed to the implementation of the method. All authors contributed to the revision, proofreading, and revision of the manuscript.

FUNDING

This work has been supported by the Fujian province Innovation Strategy Research Program (No. 2020R01020196) and YongTai Institute of Artificial Intelligence.

REFERENCES

Dolz, J., Ayed, I. B., and Desrosiers, C. (2019). "Dense multi-path U-net for ischemic stroke lesion segmentation in multiple image modalities," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Brainles 2018*, Vol. 11383, eds A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum (Cham: Springer), 271–282. doi: 10.1007/978-3-030-11723-8_27

Feigin, V. L. (2019). Anthology of stroke epidemiology in the 20th and 21st centuries: assessing the past, the present, and envisioning the future. *Int. J. Stroke* 14, 223–237. doi: 10.1177/1747493019832996

Griffis, J. C., Allendorfer, J. B., and Szaflarski, J. P. (2016). Voxel-based Gaussian naive Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans. *J. Neurosci. Methods* 257, 97–108. doi: 10.1016/j.jneumeth.2015.09.019

Haan, B. D., Clas, P., Juenger, H., Wilke, M., and Karnath, H. O. (2015). Fast semi-automated lesion demarcation in stroke. *Neuroimage Clin.* 9, 69–74. doi: 10.1016/j.nicl.2015.06.013

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 (CVPR)*, (Piscataway, NJ: IEEE), 770–778. doi: 10.1109/CVPR.2016.90

Hui, H. S., Zhang, X. Y., Li, F. L., Mei, X. B., and Guo, Y. L. (2020). A partitioning-stacking prediction fusion network based on an improved attention U-Net for stroke lesion segmentation. *IEEE Access* 8, 47419–47432. doi: 10.1109/ACCESS.2020.2977946

Ito, K. L., Kim, H., and Liew, S. L. (2019). A comparison of automated lesion segmentation approaches for chronic stroke T1-weighted MRI data. *Hum. Brain Mapp.* 40, 4669–4685. doi: 10.1002/hbm.24729

Karthik, R., Menaka, R., Johnson, A., and Anand, S. (2020). Neuroimaging and deep learning for brain stroke detection - A review of recent advancements and future prospects. *Comput. Methods Programs Biomed.* 197:105728. doi: 10.1016/j.cmpb.2020.105728

Kingma, D., and Ba, J. (2014). Adam: a method for stochastic optimization. *Comput. Sci.* arXiv [Preprint] arXiv:1412.6980.

Lee, C. Y., Xie, S., Gallagher, P., Zhang, Z., and Tu, Z. (2014). Deeply-supervised nets. *Proc. Mach. Learn. Res.* 38, 562–570. doi: 10.3390/s19092009

Liew, S.-L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., et al. (2018). A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Sci. Data* 5:180011. doi: 10.1038/sdata.2018.11

Liu, Z. Y., Cao, C., Ding, S. X., Liu, Z., Han, T., and Liu, S. (2018). Towards clinical diagnosis: automated stroke lesion segmentation on multi-spectral MR image using convolutional neural network. *IEEE Access* 6, 57006–57016. doi: 10.1109/access.2018.2872939

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, (Piscataway, NJ: IEEE), 3431–3440. doi: 10.1109/cvpr.2015.7298965

Lucas, C., Maier, O., and Heinrich, M. P. (2017). "Shallow fully-connected neural networks for ischemic stroke-lesion segmentation in MRI," in *Bildverarbeitung für die Medizin 2017*, eds Maier-Hein, K. geb. Fritzsche, Deserno, T. geb. Lehmann, H. Handels, and T. Tolxdorff (Berlin: Springer Vieweg), 261–266. doi: 10.1007/978-3-662-54345-0_59

Lyu, Z., Wang, Z., Luo, F., Shuai, J., and Huang, Y. (2021). Protein secondary structure prediction with a reductive deep learning method. *Front. Bioeng. Biotechnol.* 9:687426. doi: 10.3389/fbioe.2021.687426

Maier, O., Schroder, C., Forkert, N. D., Martinetz, T., and Handels, H. (2015). Classifiers for ischemic stroke lesion segmentation: a comparison study. *PLoS One* 10:e0145118. doi: 10.1371/journal.pone.0145118

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-Net: fully convolutional neural networks for volumetric medical image segmentation. *Med. Image Comput. Comput. Assist. Interv.* 9351, 234–241. doi: 10.1109/3dv.2016.79

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. *arXiv* [preprint]. arXiv:1804.03999,

Pandian, J. D., Gall, S. L., Kate, M. P., Silva, G. S., Akinyemi, R. O., Ovbiagele, B. I., et al. (2018). Prevention of stroke: a global perspective. *Lancet* 392, 1269–1278. doi: 10.1016/S0140-6736(18)31269-8

Park, J., Woo, S., Lee, J. Y., and Kweon, I. S. (2018). "Bam: bottleneck attention module," in *Proceedings of the British Machine Vision Conference*. arXiv [Preprint]. arXiv 1807.06514.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* 32, 8024–8035.

Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., and Avants, B. (2016). Automated segmentation of chronic stroke lesions using LINDA: lesion identification with neighborhood data analysis. *Hum. Brain Mapp.* 37, 1405–1421. doi: 10.1002/hbm.23110

Qi, K. H., Yang, H., Li, C., Liu, Z. Y., Wang, M. Y., Liu, Q. G., et al. (2019). "X-Net: brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Shenzhen, Vol. 11766, 247–255. doi: 10.1007/978-3-030-32248-9_28

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Assist. Interv.* 9351, 234–241. doi: 10.1007/978-3-319-24574-4_28

Seghier, M. L., Ramlackhansingh, A., Crinion, J., Leff, A. P., and Price, C. J. (2008). Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage* 41, 1253–1266. doi: 10.1016/j.neuroimage.2008.03.028

Tao, Y. Z., Sun, Q., Du, Q., and Liu, W. (2018). Nonlocal neural networks, nonlocal diffusion and nonlocal modeling. *Adv. Neural Inf. Process. Syst.* 31.

Woo, S. H., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "CBAM: convolutional block attention module," in *Computer Vision – ECCV 2018. Lecture Notes in Computer Science*, Vol. 11211, eds V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer), 3–19. doi: 10.1007/978-3-030-01234-2_1

Xue, Y., Farhat, F. G., Boukrina, O., Barrett, A. M., Binder, J. R., Roshan, U. W., et al. (2020). A multi-path 2.5 dimensional convolutional neural network system for segmenting stroke lesions in brain MRI images. *Neuroimage Clin.* 25:102118. doi: 10.1016/j.nicl.2019.102118

Zhang, Y., Wu, J., Liu, Y., Chen, Y., Wu, E. X., and Tang, X. (2020). MI-UNet: multi-inputs UNet incorporating brain parcellation for stroke lesion segmentation from T1-weighted magnetic resonance images. *IEEE J. Biomed. Health Inform.* 25, 526–535. doi: 10.1109/jbhi.2020.2996783

Zhang, Z., Liu, Q., and Wang, Y. (2018). Road extraction by deep residual U-Net. *IEEE Geosci. Remote Sens. Lett.* 15, 749–753. doi: 10.1109/lgrs.2018.2802944

Zhou, Y., Huang, W., Dong, P., Xia, Y., and Wang, S. (2019). "D-UNet: a dimension-fusion U shape network for chronic stroke lesion segmentation," in *Proceedings of the IEEE-ACM Transactions on Computational Biology and Bioinformatics*, Vol. 18, (Piscataway, NJ: IEEE), 940–950. doi: 10.1109/tcbb.2019.2939522

Zhu,Z., Lu, S., Wang, S.-H., Gorriz, J. M., and Zhang, Y.-D. (2021a). BCNet: a novel network for blood cell classification. *Front. Cell Dev. Biol.* 9:813996. doi: 10.3389/fcell.2021.813996

Zhu, Z., Lv, D., Zhang, X., Wang, S.-H., and Zhu, G. (2021b). Deep learning in the classification of stage of liver fibrosis in chronic hepatitis b with magnetic resonance ADC images. *Contrast Media Mol. Imaging* 2021:2015780. doi: 10.1155/2021/2015780

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

frontiers | Frontiers in Aging Neuroscience

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fnagi.2022.822480

A Network Pharmacology Based Research on the Mechanism of Donepezil in Treating Alzheimer's Disease

Lihua Liu[1†], *Yingying Zhu*[1†], *Peng Fu*[2*] *and Jundong Yang*[1*]

[1] Laizhou City People's Hospital, Laizhou, Yantai, China, [2] Department of Pharmacy, Shanghai Changhai Hospital, Naval Medical University, Shanghai, China

Objective: In order to explore and further understand the efficacy of donepezil (DNP) in the treatment of Alzheimer's disease (AD), this research was conducted based on network pharmacology and molecular docking.

Method: Compounds of DNP and its effective targets were collected using the TCMSP Chinese medicine system pharmacology database. Disease targets were screened and selected utilizing GeneCards, TTD, DrugBank, CTD, and other online databases. Then, Venn diagrams were generated to identify the intersections. A diseases-drug-active ingredient-key target protein interaction (PPI) network was constructed using the STING database. GO and KEGG enrichment analyses were conducted to predict the function and mechanism of DNP, which were visualized by graphs and bubble charts. After the screening, the top five interacting targets in the PPI network and the compound containing the most active target were selected for molecular docking.

Results: The study received 110 potential targeting genes and 155 signaling pathways. A strong association between DNP and modulation of chemical synaptic transmission and the regulation of trans-synaptic signaling is noted. Signaling pathways related to the proliferation, differentiation, and survival of cells are also found positively relative. The results revealed that the mechanism of its therapeutic effect is multi-component, multi-target, and multi-pathway, laying a foundation for the follow-up in-depth study of the mechanism of DNP in the treatment of AD.

Conclusion: This research provides a superior prediction that AD could be treated using DNP which targets the key proteins and essential pathways associated with the recovery of AD.

Keywords: donepezil, Alzheimer's disease, network pharmacolgy, mechanism, prognosis

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Hao Chen,
Hunan University, China
Luobing Dong,
Xidian University, China
Lin Haifeng,
McMaster University, Canada

***Correspondence:**
Jundong Yang
1295218809@qq.com
Peng Fu
fucaofan@163.com

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

Specialty section:
This article was submitted to Alzheimer's Disease and Related Dementias, a section of the journal Frontiers in Aging Neuroscience

Received: *25 November 2021*
Accepted: *11 February 2022*
Published: *08 April 2022*

Citation:
Liu L, Zhu Y, Fu P and Yang J (2022) A Network Pharmacology Based Research on the Mechanism of Donepezil in Treating Alzheimer's Disease. Front. Aging Neurosci. 14:822480. doi: 10.3389/fnagi.2022.822480

INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disorder in the central nervous system, prevalently observed among the elderly and near-elderly, characterized by progressive cognitive impairment, and memory degeneration (Hebert et al., 1995; Alzheimer's, 2016).

Credited as the most common cause of dementia, AD is known for its patients suffering from a long course of the disease, loss of self-care ability, and eventual death from comorbidities, identified

by the World Health Organization as a prior global public health concern (Barker et al., 2002; Wilson et al., 2012).

In 2020, an estimated number of AD cases worldwide was reported to be 24 million, which was predicted to increase fourfold by the year 2050 (Breijyeh and Karaman, 2020).

Since the first AD case was reported in 1907 by Alois Alzheimer, significant gains in our understanding of the etiology and pathogenesis of AD have been achieved. Nevertheless, current therapeutic strategies for AD remain considerably limited, with only two alternatives available, including naturally derived and synthetic inhibitors or hybrid analogs to cholinesterase enzyme, and antagonists to N-methyl D-aspartate (NMDA). Despite the effectiveness of the existing therapies in treating the symptoms of AD, no significant influence has been found in curing or preventing the disease (Wang and Reddy, 2017). Acetylcholinesterase inhibitors (AChEIs) are categorized as the former, blocking cholinesterase enzymes from degrading ACh to increase ACh levels in the synaptic cleft. For patients with all levels of AD, whether mild, moderate, or severe (Winblad et al., 2001), donepezil (DNP), an AChEI, is wildly commended in China as the most promising therapeutic candidate for AD (Deardorff and Grossberg, 2016). Studies have confirmed the efficacy and safety of DNP for patients with mild to moderate AD and its possible effectiveness for severe AD (Zhang and Gordon, 2018). However, it should also be noted that DNP is likewise only capable of treating the symptoms of AD but cannot reverse or alter the AD progression (Cacabelos, 2007).

The insufficient understanding of the pathogenesis of AD sets great obstacles in the path of finding a novel breakthrough in the improvement of the existing drugs. Therefore, the investigation of the mechanism of DNP as a drug for AD treatment should be highly prioritized.

Recently, studies concerning the mechanism of DNP are extremely limited. In 2019, Ghosh et al. revealed the mechanistic pathway of DNP-induced cholinergic inhibition of AD *via* docking and well-tempered metadynamics (WTMtD) stimulations (Duggleby et al., 2019).

METHOD

First, the structural formula of donepezil was downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/). With the structure acquired, data of the donepezil targeted sites were collected from four separate databases: swisstarget (http://www.swisstargetprediction.ch/), pharmmapper (http://lilab-ecust.cn/pharmmapper/), batman-tcm (http://bionet.ncpsb.org.cn/batman-tcm/), and stitch (http://stitch.embl.de/), each with a different standard for data selection. For swisstarget, data with Probability* >0 was included. For batman-tcm, the inclusion standard was set Score cutoff≥ 20. No standard was set for screening in stitch or pharmmapper. The corresponding genes of the donepezil target sites were collected through Uniprot (http://www.uniprot.org/www.uniprot.org/), utilizing data downloaded in pharmmapper.

Three databases were utilized in the searching of pathogenic genes of AD. In GeneCards, genes with a relevance score ≥15 were enrolled. In OMIM, 142 AD-associated genes were included with no screening conducted. In disgenet, genes with score gda ≥0.2 were included. Afterward, data from the three distinct databases were merged and had duplicates removed, leaving 819 distinct genes related to the pathogenesis of AD for subsequent research.

With the screened disease targets, intersections were taken and Venn diagrams were drawn. Four hundred and fourty nine drug targets, 819 disease targets, and 110 intersection targets were obtained. All duplications were removed.

The STRING database was used to construct a disease-drug-active ingredient-key target network, with the minimum required interaction score set as 0.4 and the isolated vertices removed. A target protein interaction network was constructed using Cytoscape3.7.2 software.

With ALB as the brightest in the string hires image and the highest in the string interaction chart, intersections related to ALB are therefore presumed to be of the highest importance.

The intersections of the core genes were collected using the R package with PPI data inputted.

Genes for potential target sites were selected using the R package from the "org.Hs.eg.db" database, and GO enrichment analysis was conducted using DOSE, clusterProfiler, and pathview package (Bioconductor) from three aspects, Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), with p-value cutoff and q-value cutoff both as 0.05. The results from each aspect were visualized with a graph and bubble chart.

With the selected genes, KEGG pathway enrichment analysis was conducted to predict its mechanism of action using DOSE, clusterProfiler, and pathview packages (Bioconductor), with p-value cutoff and q-value cutoff both as 0.05. Columns and bubble charts were generated to visualize the result.

RESULTS

Data Processing of Donepezil

To acquire the structure of donepezil, the associating structure with Probability* >0, score cutoff≥20 was screened and downloaded from the swisstarget and batman-tcm databases. The structure was also searched in the stitch and pharmmapper databases to further verify the results. The structure was visualized using PubChem (**Figure 1A**) With the selected structure, the associating genes were screened in the Uniprot database (**Table 1**).

Data Processing of AD

A total of 709, 142, and 154 associating genes are screened from the GeneCards, omim, and disgenet database, with the relevance score≥15 and score gda≥0.2. A final number of 819 genes are included after the duplicates were removed (**Table 2**). Among these genes, 449 genes are related to the drug donepezil, 819 are related to AD, and 110 intersection genes are noted. All the above research into the genes was conducted with the duplicates removed. Venny2.1 (https://bioinfogp.cnb.csic.es/tools/venny/) was utilized to visualize the intersection result (**Figure 1B** and **Table 3**).

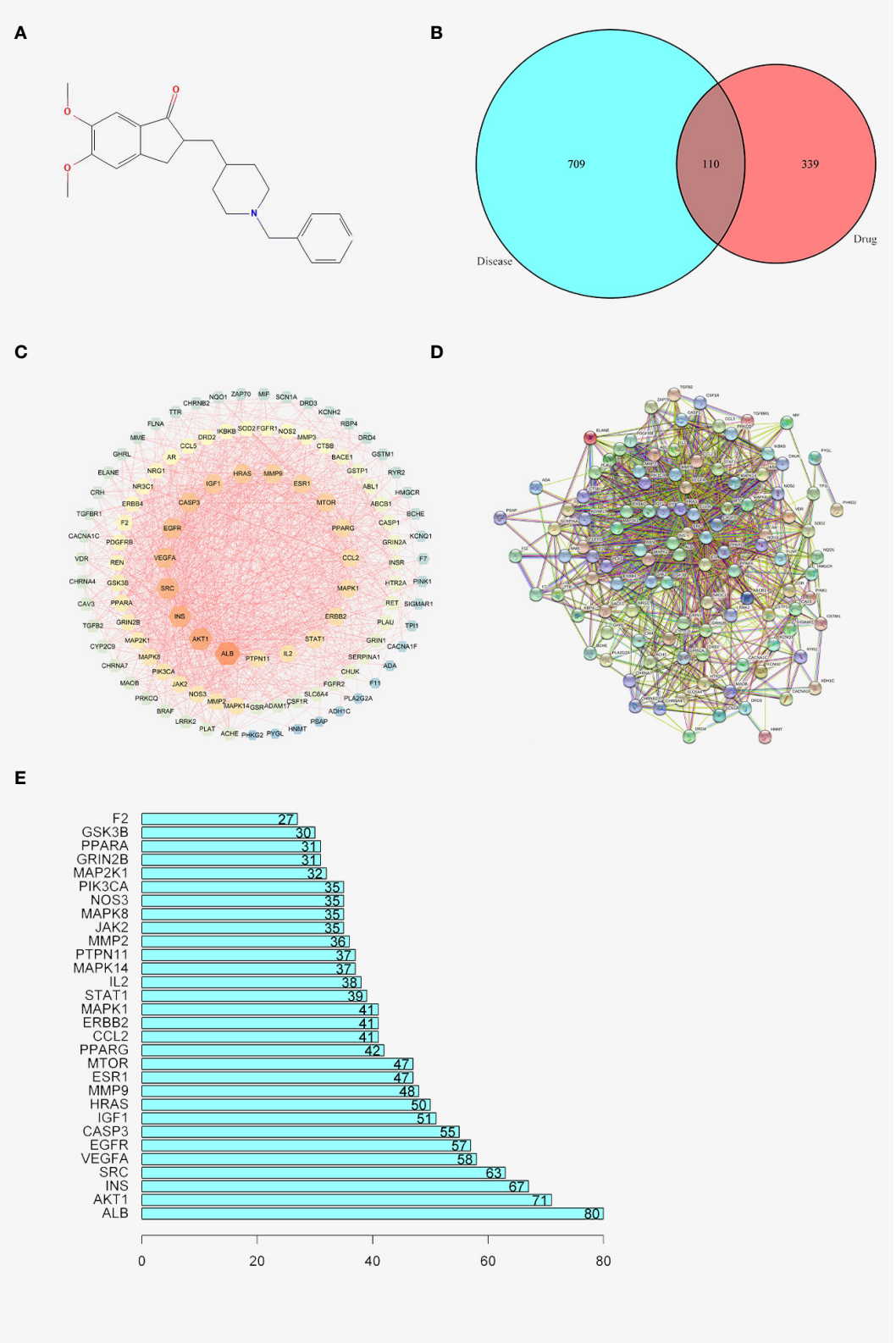


FIGURE 1 | (A) 2D structure of DNP. **(B)** Venn graph of AD and DNP. **(C)** Diseases- drug-active ingredient-key target network. **(D)** PPI network. **(E)** Bar chart of 30 core genes.

TABLE 1 | DNP associated genes.

BCHE ACHE KCNH2 HRH3 SIGMAR1 BACE1 DRD2 DRD4 HTR2A HTR2C MAPK3 MAPK1 OPRD1 OPRK1 CHEK1 MTOR PIK3CA OPRM1 MAP3K12 FGFR1 JAK3 JAK2 SCARB1 MAPK1 MAPK14 MMP2 RET ABCB1 ADRA1A GNRHR SMO AXL CHRM2 GSK3B SYK ERBB2 PAK4 BRAF PTPN11 AURKB STAT1 CHUK DRD3 RPS6KA2 HSD11B1 FLT1 LRRK2 CDK9 ADK PARP1 ABL1 KIT PDE11A JAK1 CDK4 HTR7 PLK1 PIK3CD EGLN1 CCR3 CDK2 TGFBR1 GRIN2A BDKRB1 MELK NPY1R CSF1R POLR1A FLT3 CCR4 CYP24A1 VEGFA MMP9 IGF1 PPIG CCL2 KCNB1 MMP3 ROCK1 AKR1C3 ESR1 MAPKAPK2 PIM1 PPIA PPARG CA2 CFB NR1H2 MAPK8 MAPK10 F2 ALB STS APOA2 NOS3 TTR CASP3 KIF11 EGFR MMP13 PGR AR PDE4D SRC CCNA2 ADAM17 P CACNB1 CACNA1D SCN1A CACNB3 CACNA1F ADRA1B ADRA2C ADRA2A KCNK2 CACNB2 CACNA1S ADRB1 CACNB4 ADRA1D ADRA2B SCN2B GPD1L IL2 GNB5 AIF1 FEZF2 MC2R MAP2K1 RAB3B SCN3B KCNE5 SMAD7 INS SCN4B CRH ARX FLNA CNTN2 DTNBP1 CAV3 KCNA5 PHKG2 ANK2 APLP1 CRHR1 NPPA MC5R PPP1R9B MC3R VAMP2 YWHAE FOXL2 DRD5 OPRL1 RAB3A MC1R ADCY5 AKT1 HRAS MC4R RWDD3 CAV2 ARRDC3 SYT1 DRD1 CABP1 RYR2 ADA PARK2 LHX6 EGR1 DMTN GRIN2B ADORA1 FLOT1 CALCA PINK1 KCNQ1 CASQ2 PALM GJA5 GHRL SLC18A1 PPARA

TABLE 2 | Genes related to AD.

PSEN1 APP APOE PSEN2 MAPT SNCA GBA HFE NOS3 PKD1 MT-ND1 PRNP MPO NPC1 LRRK2 TNF IGF1 IL6 IL2 GAA IL10 LMNA PRKN A2M AD5 PLAU ABCA7 MIR146A NOD2 CASP3 SMPD1 VCP AD6 AD10 MIR34A AD7 TGFB1 NEFL MFN2 ACE MIR29A AD11 AD8 AD12 VEGFA AD13 AD14 AD17 TP53 ATP7B MIR106B PARK7 ABCA1 TREM2 GDAP1 ABCA4 ADAM10 IL1B UNC5C MIR107 MTHFR PINK1 SRC ESR1 PSAP DNM2 MTOR GFAP MIR328 PMP22 AGL GDNF NPC2 BDNF GBE1 RYR1 SOD1 APBB1 HTT IFNG GARS1 PIK3CA CYBB CRP TTR MIR298 G6PC1 PTEN UCHL1 FIG4 BACE1 APOA1 GRN ALB SLC17A5 FAS INS CCL2 PON1 NAGLU MAP2K1 CTNNB1 CTSD CFTR EDNRB EGR2 SYNJ1 LRSAM1 ERBB2 SNCB CHAT TLR4 PPARG NDRG1 POLG HSD17B10 TLR2 HSPB1 EGFR KIF1B CXCL8 CYBA AKT1 APOB ACHE LAMP2 IL1A NCF2 TSC2 MIR21 IL4 MAPK1 RAB7A CLU PLP1 COMT GANAB SNCAIP LIPA GSK3B COL4A1 SOX10 LRP1 STAT3 SLC30A4 GAPDHS NFE2L2 PYY HRAS MIR505 NGB MIR766 ADAMTS1 PTGS1 MAPK1 MMP2 DNM1 APLP2 MAPK14 HSF1 IL33 CIB1 IKBKB STAT1 SERPINF1 MT2A BCL2L2 INS-IGF2 ADAM9 CASP12 FERMT2 PTK2B PLCB1 PPARA GRIN2B JAK2 PTPN11

TABLE 3 | Intersection genes between AD and DNP related genes.

BCHE ACHE KCNH2 SIGMAR1 BACE1 DRD2 DRD4 HTR2A MAPK1 MTOR PIK3CA SLC6A4 PDGFRB FGFR1 JAK2 PRKCQ RET ABCB1 ERBB2 BRAF IKBKB CHUK DRD3 LRRK2 ABL1 PDE11A TGFBR1 GRIN2A CSF1R VEGFA MMP9 IGF1 CCL2 MMP3 ESR1 MAPK8 F2 ALB TTR CASP3 AR SRC ADAM17 MAPK14 NOS3 NQO1 PPARG MIF ADH1C ERBB4 PTPN11 GSK3B PLAU F7 FGFR2 PLAT GSTP1 PPARA CYP2C9 REN SOD2 MAOB RBP4 MMP2 SERPINA1 ELANE EGFR PLA2G2A TPI1 MME VDR PSAP ZAP70 IL2 HMGCR HNMT F11 NR3C1 PYGL TGFB2 STAT1 CCL5 MAP2K1 CTSB GSTM1 NOS2 CASP1 GSR INSR AKT1 HRAS CHRNB2 GRIN2B CHRNA4 CHRNA7 GRIN1 CACNA1C NRG1 SCN1A CACNA1F INS CRH FLNA CAV3 PHKG2 RYR2 ADA PINK1 KCNQ1 GHRL

Construction and Analysis of Disease-Drug-Active Ingredient-Key Target Network

The STRING database was used to construct a disease-drug-active ingredient-key target network, with the minimum required interaction score set as 0.4 and the isolated vertices removed (**Figure 1C**). The brightness, the bigger the dots are, the higher the degree is, representing the importance of the spot to the network.

TABLE 4 | 109 core genes and their interaction.

Order	Gene	Association
1	ALB	80
2	AKT1	71
3	INS	67
4	SRC	63
5	VEGFA	58
6	EGFR	57
7	CASP3	55
8	IGF1	51
9	HRAS	50
10	MMP9	48
11	ESR1	47
12	MTOR	47
13	PPARG	42
14	CCL2	41
15	ERBB2	41
16	MAPK1	41
17	STAT1	39
18	IL2	38
19	MAPK14	37
20	PTPN11	37
21	MMP2	36
22	JAK2	35
23	MAPK8	35
24	NOS3	35
25	PIK3CA	35
26	MAP2K1	32
27	GRIN2B	31
28	PPARA	31
29	GSK3B	30
30	F2	27

Construction and Analysis of PPI Network

To further investigate the mechanism of donepezil's therapeutic effect on AD, 110 intersecting genes were enrolled into the STRING database for PPI network analysis. One hundred and nine core genes and their interacting lines were screened and visualized (**Figure 1D** and **Table 4**). The intersections of the core genes were collected using the R package with inputted PPI data. Thirty core genes were finally screened and visualized (**Figure 1E**).

GO and KEGG Enrichment Analysis

Genes for potential target sites were selected using the R package from "org.Hs.eg.db" database, and GO enrichment analysis was conducted using the DOSE, clusterProfiler, and pathview packages (Bioconductor) from three aspects, Biological

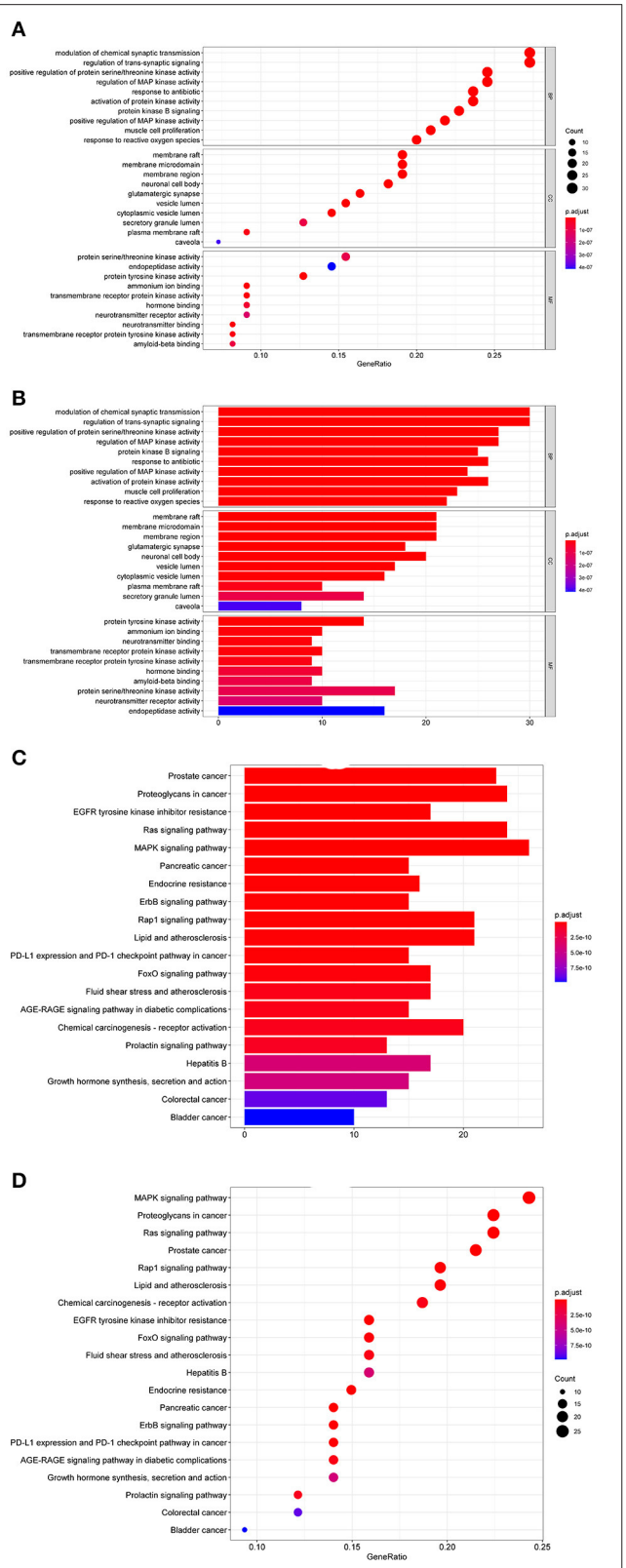


FIGURE 2 | (A) Bubble graph of GO enrichment analysis. **(B)** Bar chart of GO enrichment analysis. **(C)** Bar chart of KEGG enrichment analysis. **(D)** Bubble graph of KEGG enrichment analysis.

Process (BP), Cellular Component (CC), and Molecular Function (MF), with p-value cutoff and q-value cutoff both as 0.05. The results from each aspect were visualized with a bubble and bar chart (**Figures 2A,B**). 30 GO lists are acquired ($P < 0.05$, FDR<0.05), including 10 BP, 10CC, and 10 MF. Therefore, it can be concluded that the mechanism of the therapeutic effect of donepezill to AD could be associated with biological processes such as positive regulation of RNA polymerase II promoter, signal transduction, and positive regulation of DNA template transcription. At the same time, a variety of substances such as plasma membrane, cytosol, and extracellular space are involved, such as protein binding, enzyme binding, transcription factor activity, sequence-specific DNA binding, same protein binding, and other molecular functions.

In KEGG pathway enrichment analysis, 155 signaling pathways are screened ($P < 0.05$, FDR<0.05), and the top 20 are listed by Count number (**Figure 2C**). The KEGG analysis indicates that the genes are especially enriched in the MAPK signaling pathway, Ras signaling pathway, and Proteoglycan in cancer. An advanced bubble graph is formed based on the above results, with the signaling pathways as the ordinate and the gene ratio as the abscissa. The count value represents the number of targets enriched on the signal path. The larger the bubble, the more targets enriched, and the color of the bubble represents the p-value (**Figure 2D**).

DISCUSSION

The development of drugs with clinical value has met with extreme obstacles (Becker et al., 2008). Over the past 50 years, only two types of drugs are approved for the treatment of AD, which are cholinesterase inhibitors and memantine. Approximately 200 drugs were advanced to at least phase 2 development. However, the therapeutic effects of the approved drugs are limited and there have been disputes on whether they are clinically useful (Schneider et al., 2014).

DNP is a kind of AChEI, which functions as an inhibitor to the cholinesterase enzyme and sustains the activity of acetylcholine at cholinergic synapses. DNP has been shown to significantly improve cognition and daily function and some behavior manifestations of patients with AD (Massoud and Gauthier, 2010). In this study, DNP was taken as the research object. By using the methods of network pharmacology and molecular docking, through the establishment of a disease-drug-active ingredient-key target network, the possible targets and signaling pathways in the treatment of AD were analyzed. The results revealed that the mechanism of its therapeutic effect is multi-component, multi-target, and multi-pathway, laying a foundation for the follow-up in-depth study of the mechanism of DNP in the treatment of AD.

The study received 110 potential targeting genes and 155 signaling pathways in total. The diversity of the associated genes and pathways demonstrate that the mechanism of DNP in the treatment of AD is multifactorial. A strong association between DNP and modulation of chemical synaptic transmission and the regulation of trans-synaptic signaling is noted, further

confirming the ability of DNP in sustaining the activity of synapses. In addition, signaling pathway related to the proliferation, differentiation, and survival of cells are also found positively relative, such as Ras and MAPK signaling pathways, suggesting a potential of promoting the growth and proliferation of neurons using DNP.

The cholinergic system plays an indispensable role in neuronal function in memory, learning, and other essential aspects of cognition and plays a wider role in the promotion of neuronal plasticity (Drachman and Leavitt, 1974). Human studies of AD have shown that lesion of the cholinergic system, emerging as early as asymptomatic or prodromal stages of the disease, is one of the most important factors (Schroder et al., 1991). It has been discovered that promotion of cholinergic function in patients with AD may also have durable beneficial biological effects on the brain besides a temporary augmentation of cognitive function (Cavedo et al., 2016, 2017). Cholinesterase inhibitors are designed to inhibit the acetylcholine breakdown and sustain its activity at synapses. There has been evidence that DNP can restore cholinergic function by blocking the enzymes that break down acetylcholine (Lovestone and Howard, 1995). DNP belongs to the most currently available FDA-approved cholinesterase inhibitors for the treatment of AD (Hogan, 2007; Massoud and Gauthier, 2010). Since the number of patients with AD has been on the rise in recent decades, and is becoming a severe social concern, the need to develop an effective drug or therapeutic management for AD is urgent. It is indicated that the implication of cholinesterase inhibitor significantly reduces the risk for nursing home placement by ∼30% annually (Feldman et al., 2009). In future studies, more attention should be paid on the mechanism of the pathogenesis of AD and the mechanism of DNP for AD treatment.

In this study, due to the incomplete information in the database and unclear interaction between DNP and other factors, the prediction results have certain limitations. This study did not verify the possible molecular mechanism of DNP in the treatment of AD syndromes either *in vivo* or *in vitro*, and deficiencies remain noticeable. However, this study can provide new ideas and directions for further exploration of relevant experiments.

DATA AVAILABILITY STATEMENT

The simulation experiment data used to support the findings of this study are available from the corresponding author upon request.

AUTHOR CONTRIBUTIONS

LL and PF prepared the figures. JY and YZ finished the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the Science and Technology Commission of Shanghai Municipality, General Program, No.20ZR1456400.

REFERENCES

Alzheimer's (2016). Alzheimer's disease facts and figures. *Alzheimers Dement*. 12, 459–509. doi: 10.1016/j.jalz.2016.03.001

Barker, W. W., Luis, C. A., Kashuba, A., Luis, M., Harwood, D. G., Loewenstein, D., et al. (2002). Relative frequencies of Alzheimer disease, lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer. Dis. Assoc. Disord.* 16, 203–212. doi: 10.1097/00002093-200210000-00001

Becker, R. E., Greig, N. H., and Giacobini, E. (2008). Why do so many drugs for Alzheimer's disease fail in development? Time for new methods and new practices? *J. Alzheimers Dis.* 15, 303–325. doi: 10.3233/JAD-2008-15213

Breijyeh, Z., and Karaman, R. (2020). Comprehensive review on Alzheimer's disease: causes and treatment. *Molecules* 25:5789. doi: 10.3390/molecules25245789

Cacabelos, R. (2007). Donepezill in Alzheimer's disease: from conventional trials to pharmacogenetics. *Neuropsychiatr. Dis. Treat.* 3, 303–333.

Cavedo, E., Dubois, B., Colliot, O., Lista, S., Croisile, B., Tisserand, G. L., et al. (2016). Reduced regional cortical thickness rate of change in donepezill-treated subjects with suspected prodromal Alzheimer's disease. *J. Clin. Psychiatry* 77:e1631–e1638. doi: 10.4088/JCP.15m10413

Cavedo, E., Grothe, M. J., Colliot, O., Lista, S., Chupin, M., Dormont, D., et al. (2017). Reduced basal forebrain atrophy progression in a randomized donepezill trial in prodromal Alzheimer's disease. *Sci. Rep.* 7:11706. doi: 10.1038/s41598-017-09780-3

Deardorff, W. J., and Grossberg, G. T. (2016). Fixed-dose combination of memantine extended-release and donepezill in the treatment of moderate-to-severe Alzheimer's disease. *Drug Des. Devel. Ther.* 10, 3267–3279. doi: 10.2147/DDDT.S86463

Drachman, A., and Leavitt, J. (1974). Human memory and the cholinergic system. A relationship to aging? *Arch. Neurol.* 30, 113–121. doi: 10.1001/archneur.1974.00490320001001

Duggleby, W., Ploeg, J., McAiney, C., Fisher, K., Jovel Ruiz, K., Ghosh, S., et al. (2019). A comparison of users and nonusers of a web-based intervention for carers of older persons with alzheimer disease and related dementias: mixed methods secondary analysis. *J. Med. Internet. Res.* 21:e14254. doi: 10.2196/14254

Feldman, H. H., Pirttila, T., Dartigues, J. F., Everitt, B., Van Baelen, B., Schwalen, S., et al. (2009). Treatment with galantamine and time to nursing home placement in Alzheimer's disease patients with and without cerebrovascular disease. *Int. J. Geriatr. Psychiatry* 24, 479–488. doi: 10.1002/gps.2141

Hebert, L. E., Scherr, P. A., Beckett, L. A., Albert, M. S., Pilgrim, D. M., Chown, M. J., et al. (1995). Age-specific incidence of Alzheimer's disease in a community population. *JAMA* 273, 1354–1359. doi: 10.1001/jama.1995.03520410048025

Hogan, D. B. (2007). Progress update: pharmacological treatment of Alzheimer's disease. *Neuropsychiatr. Dis. Treat.* 3, 569–578.

Lovestone, S., and Howard, R. (1995). Alzheimer's disease: a treatment in sight? *J. Neurol. Neurosurg. Psychiatry* 59, 566–567. doi: 10.1136/jnnp.59.6.566

Massoud, F., and Gauthier, S. (2010). Update on the pharmacological treatment of Alzheimer's disease. *Curr. Neuropharmacol.* 8, 69–80. doi: 10.2174/157015910790909520

Schneider, L. S., Mangialasche, F., Andreasen, N., Feldman, H., Giacobini, E., Jones, R. et al. (2014). Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. *J. Intern. Med.* 275, 251–283. doi: 10.1111/joim.12191

Schroder, H., Giacobini, E., Struble, R. G., Zilles, K., and Maelicke, A. (1991). Nicotinic cholinoceptive neurons of the frontal cortex are reduced in Alzheimer's disease. *Neurobiol. Aging*. 12, 259–262. doi: 10.1016/0197-4580(91)90107-U

Wang, R., and Reddy, P. H. (2017). Role of glutamate and NMDA receptors in alzheimer's disease. *J. Alzheimers Dis*. 57, 1041–1048. doi: 10.3233/JAD-160763

Wilson, R. S., Segawa, E., Boyle, P. A., Anagnos, S. E., Hizel, L. P., Bennett, D. A., et al. (2012). The natural history of cognitive decline in Alzheimer's disease. *Psychol. Aging*. 27, 1008–1017. doi: 10.1037/a0029857

Winblad, B., Engedal, K., Soininen, H., Verhey, F., Waldemar, G., Wimo, A., et al. (2001). A 1-year, randomized, placebo-controlled study of donepezill in patients with mild to moderate AD. *Neurology* 57, 489–495. doi: 10.1212/WNL.57.3.489

Zhang, N., and Gordon, M. L. (2018). Clinical efficacy and safety of donepezill in the treatment of Alzheimer's disease in Chinese patients. *Clin. Interv. Aging* 13, 1963–1970. doi: 10.2147/CIA.S159920

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

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

Copyright © 2022 Liu, Zhu, Fu 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 Neuroscience

ORIGINAL RESEARCH
published: 19 April 2022
doi: 10.3389/fnins.2022.870520

High-Resolution Digital Panorama of Multiple Structures in Whole Brain of Alzheimer's Disease Mice

Xianzhen Yin[1,2,3*†], *Xiaochuan Zhang*[1,2†], *Jingjing Zhang*[2†], *Weicheng Yang*[4], *Xian Sun*[1], *Haiyan Zhang*[2*], *Zhaobing Gao*[2,5*] and *Hualiang Jiang*[2,3,4,6*]

[1] Center for MOST and Image Fusion Analysis, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China, [2] CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China, [3] Lingang Laboratory, Shanghai, China, [4] School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing, China, [5] Zhongshan Institute of Drug Discovery, Institution for Drug Discovery Innovation, Chinese Academy of Science, Zhongshan, China, [6] School of Life Science and Technology, Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
James C. Vickers,
University of Tasmania, Australia
Thomas Enzlein,
Mannheim University of Applied Sciences, Germany

***Correspondence:**
Xianzhen Yin
xzyin@simm.ac.cn
Haiyan Zhang
hzhang@simm.ac.cn
Zhaobing Gao
zbgao@simm.ac.cn
Hualiang Jiang
hljiang@simm.ac.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 06 February 2022
Accepted: 11 March 2022
Published: 19 April 2022

Citation:
Yin X, Zhang X, Zhang J, Yang W, Sun X, Zhang H, Gao Z and Jiang H (2022) High-Resolution Digital Panorama of Multiple Structures in Whole Brain of Alzheimer's Disease Mice. Front. Neurosci. 16:870520. doi: 10.3389/fnins.2022.870520

Simultaneously visualizing Amyloid-β (Aβ) plaque with its surrounding brain structures at the subcellular level in the intact brain is essential for understanding the complex pathology of Alzheimer's disease, but is still rarely achieved due to the technical limitations. Combining the micro-optical sectioning tomography (MOST) system, whole-brain Nissl staining, and customized image processing workflow, we generated a whole-brain panorama of Alzheimer's disease mice without specific labeling. The workflow employed the steps that include virtual channel splitting, feature enhancement, iso-surface rendering, direct volume rendering, and feature fusion to extract and reconstruct the different signals with distinct gray values and morphologies. Taking advantage of this workflow, we found that the denser-distribution areas of Aβ plaques appeared with relatively more somata and smaller vessels, but show a dissimilar distributing pattern with nerve tracts. In addition, the entorhinal cortex and adjacent subiculum regions present the highest density and biggest diameter of plaques. The neuronal processes in the vicinity of these Aβ plaques showed significant structural alternation such as bending or abrupt branch ending. The capillaries inside or adjacent to the plaques were observed with abundant distorted micro-vessels and abrupt ending. Depicting Aβ plaques, somata, nerve processes and tracts, and blood vessels simultaneously, this panorama enables us for the first time, to analyze how the Aβ plaques interact with capillaries, somata, and processes at a submicron resolution of 3D whole-brain scale, which reveals potential pathological effects of Aβ plaques from a new cross-scale view. Our approach opens a door to routine systematic studies of complex interactions among brain components in mouse models of Alzheimer's disease.

Keywords: simultaneous visualization, multiple structures, Alzheimer's disease, micro-optical sectioning tomography, cross-scale

INTRODUCTION

Amyloid-β (Aβ) plaque deposition is one major pathological hallmark of Alzheimer's disease (AD) (Murphy and LeVine, 2010). Aβ plaques are surrounded by multiple brain structures such as neurons, glia cells, nerve fibers, and blood vessels in the AD brain (Marin et al., 2016; Yuan et al., 2016; Miners et al., 2018; Greenberg et al., 2020). The structural abnormalities of these structures have been observed in the AD brains, which include cerebral amyloid angiopathy (Gurol et al., 2016), white matter (WM) lesion (Bozzali et al., 2002), ventricular enlargement (Apostolova et al., 2012), and brain atrophy (Pini et al., 2016). In addition, Aβ plaques-associated dystrophic axons (Adalbert et al., 2009), dendritic spine loss and synaptic alterations (Bittner et al., 2012), separation of astrocyte end-feet from the endothelial vessel wall (Kimbrough et al., 2015), recruitment and activation of glial cells (Fakhoury, 2018), and neuron death (Kadowaki et al., 2005) were also identified. These findings highlighted the essential contributions of in-depth analyzing the interactions of Aβ plaque with its surroundings to a thorough understanding of the pathological mechanism and progress of AD. However, the simultaneous visualization that shows how the plaques affect their surrounded neurons, nerve processes and tracts, and blood vessels in high resolution and whole brain is rarely achieved.

A number of two categories of methods have been developed and wildly used to image Aβ plaques. The labeling methods require specific reagents binding to aggregated Aβ, which include amyloid binding dyes thioflavin S or T (Bacskai et al., 2003), fluorochrome methoxy XO4 (Klunk et al., 2002), and Aβ-specific antibodies (Youmans et al., 2012), and *in vivo* near-infrared fluorescent probes such as DANIR 8c (Fu et al., 2016). In contrast, the label-free methods, such as stimulated Raman scattering microscopy (Ji et al., 2018), optical coherence microscopy (Bolmont et al., 2012), autofluorescence under liquid nitrogen freezing (Luo et al., 2017), multiphoton excitation fluorescence and second harmonic generation (Johansson and Koelsch, 2017), and MALDI mass spectrometry imaging (Enzlein et al., 2020), could avoid potential interference induced by exogenous molecules and provide more options for Aβ plaque imaging. Noticeably, the emerging of efficient brain-wide optical imaging techniques coupled with fast-developing amyloid labeling methods allows whole-brain imaging of Aβ plaques. Combining iDISCO, a novel clearing method integrating whole-mount antibody immunolabeling, the light-sheet microscopy permitted 3D detection of amyloid pattern in a full mouse brain hemisphere and archived human brain tissues (Liebmann et al., 2016). Fluorescence micro-optical sectioning tomography (fMOST) was successfully used to detect the brain-wide distribution of DANIR 8c labeled Aβ plaques in mice (Long et al., 2019). Similarly, serial two-photon tomography imaging of methoxy-X04 labeled plaques was applied to obtain the whole-brain imaging of amyloid pathology in AD mice (Whitesell et al., 2019). As to label-free brain-wide imaging of Aβ plaques, cryo-micro-optical sectioning tomography (cryo-MOST) was developed to acquire the whole-brain coronal distribution of Aβ plaques using intrinsic fluorescence emission under liquid nitrogen freezing (Luo et al., 2017).

To image brain vasculature, some new techniques drew more attention due to enhanced spatial resolution or extensive potential applications in comparison with the classic methods such as computed tomography angiography (Leiva-Salinas et al., 2018) and magnetic resonance angiography (Dorr et al., 2007). For example, the vascular corrosion cast coupled with stereomicroscopic scanning was utilized to visualize the 3D arrangement of the entire brain vasculature in APP23 mice (Meyer et al., 2008). Serial two-photon microscopy (Delafontaine-Martel et al., 2018) and light sheet microscopy (Di Giovanna et al., 2018) combined with intracardiac perfusion of FITC-gelatin perfusate achieved whole-brain vascular imaging at high resolution. Except the FITC-gelatin gel perfusion method, whole-mount immunolabeling (Kirst et al., 2020), fluorescent tagged transgenic mice (Jing et al., 2018), and conventional vascular-labeling dyes such as fluorescent lectin (Khouri et al., 2021) or lipophilic tracers (Konno et al., 2020) were also been adopted in light sheet microscopy technique for whole-brain vascular imaging. Interestingly, a novel method named "SeeNet" integrating multiple technologies (vascular casting with intravascular perfusion of fluorescent cross-linker, tissue clearing, and light sheet microscopy) achieved whole-brain visualization of cerebral vasculatures at the single-micro-vessel level (Miyawaki et al., 2020). Noticeably, the MOST/fMOST exhibited powerful capability to describe the entire vasculature of the mouse brain at high resolution with or without fluorescence labeling (Gang et al., 2017; Xiong et al., 2017). In one of our recent studies, using MOST and modified Nissl staining, we generated a cross-scale whole-brain 3D vascular atlases of AD mice, which covers the entire vascular system from large vessels down to the smallest capillaries at submicron resolution (Zhang et al., 2019).

Combined with the specific labeling, the whole-brain imaging techniques were also applicable to image and analyze specific-function neurons and neural circuits. *In vivo* fluorescence labeling such as intravascular injections of fluorescent dyes (Appaix et al., 2012), intracerebral injection of neurotropic viruses (Zingg et al., 2017), and varieties of fluorescent transgenic mice (Daigle et al., 2018) allows systematic imaging of molecularly labeled structures within the large intact brain. Serial two-photon tomography permitted collecting high-resolution datasets of fluorescent labeled Thy1-GFP mouse brains and also imaging adeno-associated virus (AAV)-injected brains expressing GFP for anterograde tracing (Ragan et al., 2012). Light sheet microscopy integrating tissue clearing and virus tracers or transgenic mice was widely used in mapping neural circuits (Jing et al., 2018; Mano et al., 2018). Lattice light-sheet microscopy coupled with expanded Thy1-GFP mouse brain realized nanoscale brain-wide optical imaging of neural circuits (Gao et al., 2019). Using the fMOST technique and multiple labeling, several brain research projects have sought to reconstruct high-resolution neuronal morphology of regions of interest and to provide the visualizations of neural pathways (Yuan et al., 2015).

These abovementioned advanced imaging methods permit high-resolution volume imaging of Aβ plaques, neurons, neurites, or blood vessels, individually. To study the complex influences of Aβ plaques in the AD brain, visualizing structures

as many as possible in one same sample simultaneously will be greatly helpful. Due to the spectral overlap of the fluorophore and crossover of fluorescence emission, it is still of huge challenges to detect more than three fluorescent labels using fluorescence imaging techniques (McRae et al., 2019). Employing the reflected bright-field imaging technique of MOST with a whole-brain Nissl staining method, our previous study demonstrated the capability to distinguish different structures that include somata, blood vessels, and nerve processes based on the differences in gray values and morphologies (Zhang et al., 2019). In this study, taking advantage of the powerful capability of MOST in dissecting gray values and a newly developed customized image processing workflow for extraction and reconstruction of multiple structural signals, we successfully visualized the Aβ plaques, nerve tracts and nerve processes, somata, and blood vessels simultaneously in the whole brain of AD mice for the first time, without any specific labeling. This research provides a novel approach to clearly present various structural information in a whole mouse brain simultaneously, which will facilitate a better understanding of the cerebral anatomical features under the pathological state of AD.

RESULTS

Acquisition of High-Content Characteristic Signals in Same Whole Mouse Brain

The high-content whole-brain datasets were obtained by combining the MOST system and whole-brain Nissl staining method. To acquire high contrast images, the bright-field MOST imaging was adjusted to a stable state with high image contrast during the time-consuming acquisition. Representative images of the coronal section of the whole-brain Nissl staining datasets in 6-month-old wild-type (WT) and 5×FAD mice (co-expressing five mutations of familial Alzheimer's disease) are shown in **Figures 1A,B**. The enlarged view in **Figures 1C,D** showed the different morphological features of multiple brain structures that include blood vessels (red arrow), nerve processes (blue arrow), and somata (purple arrow). The blood vessels were presented in a higher gray value comparing with surrounding somata and parenchyma, which makes them being easily visualized (**Figures 1E,F**). The signal of nerve processes was particularly evident in the strata radiatum of the hippocampus, with a gray value between that of vessels and somata (**Figures 1C–F**).

These abovementioned structures were observed with similar gray-level distribution in both WT and 5×FAD mice. Noticeably, a large amount of pompon-like structures, which obviously differed from other neuroanatomical brain structures, were observed in 5×FAD mice (cyan arrow in **Figure 1D**) rather than WT mice. Most of these pompon-like structures had a dark core surrounded by paler regions. Correspondingly, the gray-intensity plot of the pompons showed a specific "W" shape, owing to the relatively lower gray value in the center and the higher gray value in the rim (**Figure 1F**). In addition to these abovementioned signals, the massive dark circles in the striatum could also be identified according to their low gray level, large diameter, and distinctive morphology. As shown in **Figures 1G,H**, the massive dark circles in the striatum were supposed to be nerve tract. In conclusion, multiple discernible signals of different structures were simultaneously acquired using the whole-brain Nissl staining and MOST system.

Customized Image Processing Workflow for Identification and Extraction of Multiple Structures

Although the characteristic signals of somata, blood vessels, and nerve processes and tracts had been acquired, the visual rendering of these multiple structures encountered considerable obstacles due to the heterogeneous gray levels and low contrast of the bright field images. To identify and validate exactly what the abovementioned characteristic signals are, a customized image processing workflow for extraction and reconstruction of these multiple structures was developed (**Supplementary Figure 1**). Initially, the coronal images were optimized to enhance image quality through the steps that include background correction, noise reduction, and contrast enhancement, which had been described in our previous study (Zhang et al., 2019). The enhanced contrast of the optimized high-quality coronal images made it more feasible to identify the ultrastructure of these structures based on the grayscale and morphological differences. The vessels in the highest gray ranges and the somata in the lowest gray ranges could be easily visualized. In addition, the information of processes, tracts, and plaques was difficult to be extracted due to their adjacent gray-value distribution. Therefore, the virtual channel splitting method was performed to individually extract the characteristic structure information of blood vessels, pompons, and neural structures. The vascular information was filtered out by region growth algorithm and then was reconstructed using iso-surface rendering; the pompon information was visualized by high-precision real-time rendering after feature enhancement; the neural information that includes soma, nerve process, and tract was extracted and then performed direct volume rendering. At last, all the separated virtual channels were integrated through feature fusion to complete synchronous visualization of multiple structures.

After the extraction and reconstruction of 3D architectures, the characteristic signals of neural structures and blood vessels were identified (**Supplementary Figure 1**) and were also confirmed by referring to the observed structures in the other studies (Wu et al., 2014; Xiong et al., 2017). The pompons observed in this study could not be classified as any reported neural structures, but resembled the morphology of Aβ plaques observed in other imaging studies using Thioflavin S staining (Dickson and Vickers, 2001). To examine whether the pompon-like structures were Aβ plaques, we performed whole-brain Nissl staining, thioflavin S (ThS) staining, and anti-Aβ antibody labeling on the same paraffin tissue sections for comparison. The large-size images of the coronal brain section from a 6-month-old 5×FAD mouse are shown in **Figures 2A–D**. Typical brain architectures were easily recognized in all images, which include the cortex, hippocampal cornu ammonis, and dentate gyrus. The enlarged

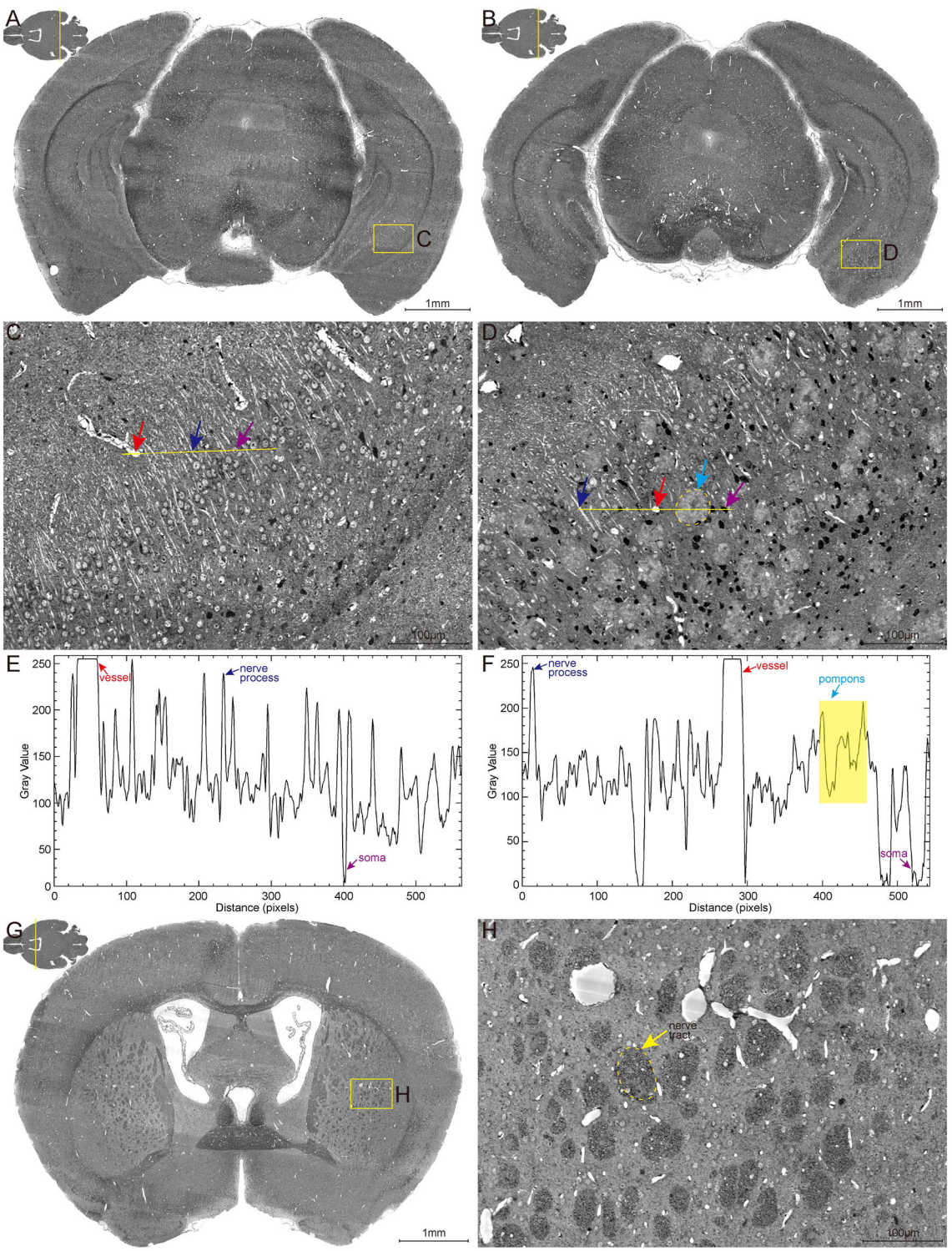


FIGURE 1 | High-resolution brain atlas showing multiple signals of pompons, somata, nerve processes and tracts, and blood vessels. **(A,C,E)** WT group; **(B,D,F)** 5×FAD group. **(A,B)** Representative images of hippocampal coronal section in the 6-month-old WT and 5×FAD mice. **(C,D)** Enlarged views of the boxes in **(A,B)** show the multiple signals in Nissl staining data. Red, blue, purple, and cyan arrows denote vessels, nerve processes, somata, and pompons, respectively. **(E,F)** Two-dimensional graphs of the gray intensities of pixels along the drawn line in **(C,D)**. The x-axis represents distance along the line, and the y-axis is the gray intensity. The leftmost end of lines was corresponding to the leftmost end of the plot profile. Red, blue, purple, and cyan arrows denote the distinctive gray value of the representative vessels, nerve processes, somata, and pompons marked in **(C,D)**. **(G,H)** Representative image of coronal section across striatum in 5×FAD mouse shows the apparent signal of nerve tract. Enlarged views of the box in **(G)** are shown in **(H)**. The dotted box indicated by the yellow arrow denotes the nerve tract.

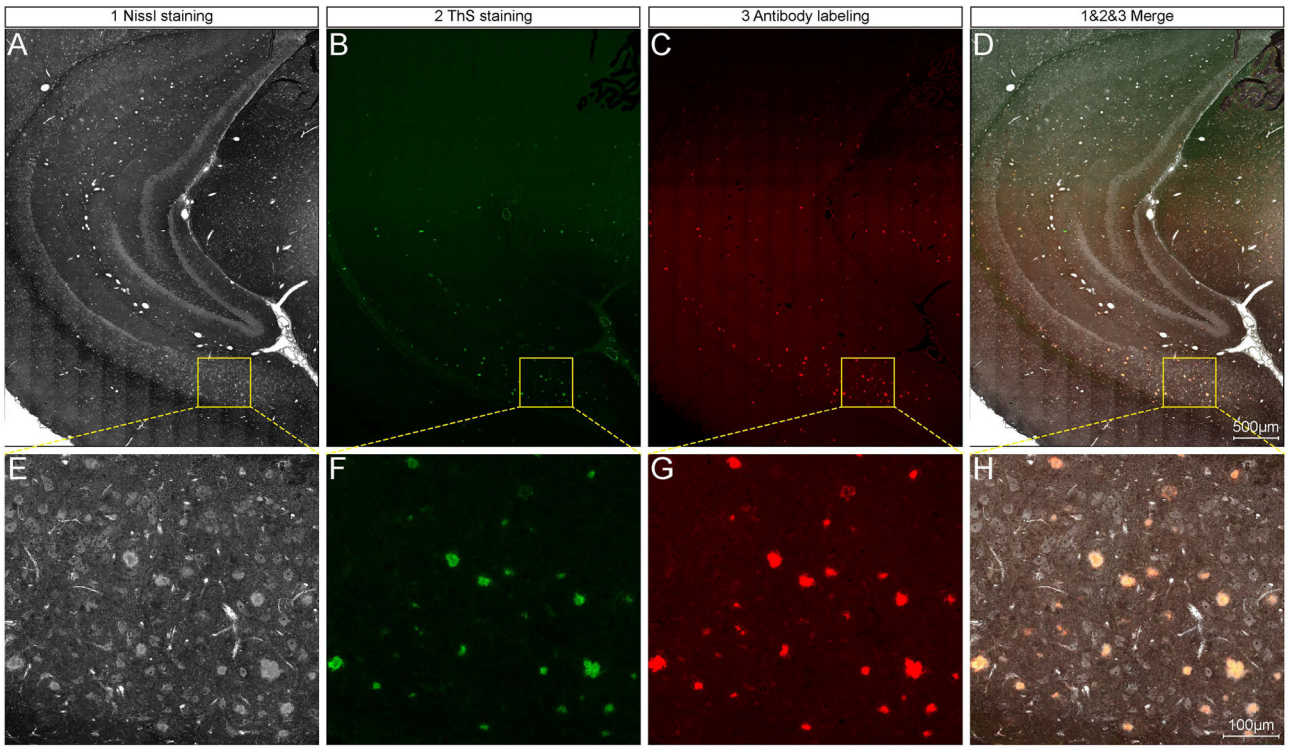


FIGURE 2 | Comparison between whole-brain Nissl staining and Aβ-specific staining results on the same tissue section. **(A–C)** Representative images showing plaque signals identified by whole-brain Nissl staining **(A)**, ThS staining **(B)**, and antibody labeling **(C)** on the same tissue section of 6-month-old 5×FAD mice. **(D)** Overlapping of images in **(A–C)**. Enlarged views of the boxes in **(A–D)** are corresponding shown in **(E–H)**.

views of yellow boxes in **Figures 2A–D** were displayed in **Figures 2E,H**, respectively. From the representative images of the Nissl staining coronal section in **Figures 2A,E**, the pompon-like structures can be clearly distinguished. Both the ThS fluorescence (**Figures 2B,F**) and antibody immunofluorescence (**Figures 2C,G**) were observed on the same coronal section as shown in **Figures 2A,E**. In regions where Aβ plaques were detected by ThS and anti-Aβ antibody, whole-brain Nissl staining could also successfully label them with one-to-one correlation. The merged images in **Figures 2D,H** illustrated the co-localization of Nissl, ThS, and anti-Aβ antibody labeled Aβ plaques. These results demonstrated that the pompon-like structures obtained by the whole-brain Nissl staining method were indeed Aβ plaques.

Generation of All-in-One Panorama for Aβ Plaques, Somata, Nerve Processes and Tracts, and Blood Vessels in Whole Mouse Brain

Brain-Wide Visualization of Aβ Plaque Distributions

Systematic whole-brain visualization of Aβ plaque distribution is critical to characterize the key pathological features for understanding the pathology of AD. **Figure 3A** and **Supplementary Movie 1** showed the distribution of Aβ plaques across an entire brain of a 5×FAD mouse. The brain-wide plaque distribution from the olfactory bulb to cerebellum was clearly visualized, and the plaques were observed throughout the whole brain. A 3D heatmap of whole-brain plaque distribution density was shown in **Figure 3B** and **Supplementary Movie 2**. The deeper layer of the cortex exhibited a denser plaque distribution, and the entorhinal cortex along with its adjacent subiculum regions presented the highest density of plaques. To confirm the plaque distribution, we compared the plaque numbers in the entorhinal cortex and subiculum regions with those in their neighboring regions. **Table 1** showed that the plaque numbers in the entorhinal cortex and subiculum regions were much larger than those in the adjacent cortex and hippocampal CA3 regions, which was consistent with the results from the 3D heatmap. Quantification analysis of the whole brain plaque number in different diameters revealed that the Aβ plaques smaller than 40 μm accounted for the majority of total plaques (**Figure 3C**). In addition, a top view of the brain-wide plaque distribution and its corresponding diameter-coded colormap were displayed in **Figures 3D,E**, respectively. The plaque distribution pattern was analyzed on three diameter ranges: diameter $< 30\,\mu m$, $30\,\mu m <$ diameter $< 40\,\mu m$, diameter $> 40\,\mu m$ (**Figures 3F–H**). In accordance with the quantification analysis in **Figure 3C**, the plaque number in these three ranges of diameter exhibited huge differences. The Aβ plaques smaller than 30 μm in diameter were the most plentiful, and there was no apparent difference in plaque density across the whole brain (**Figure 3F**). The Aβ

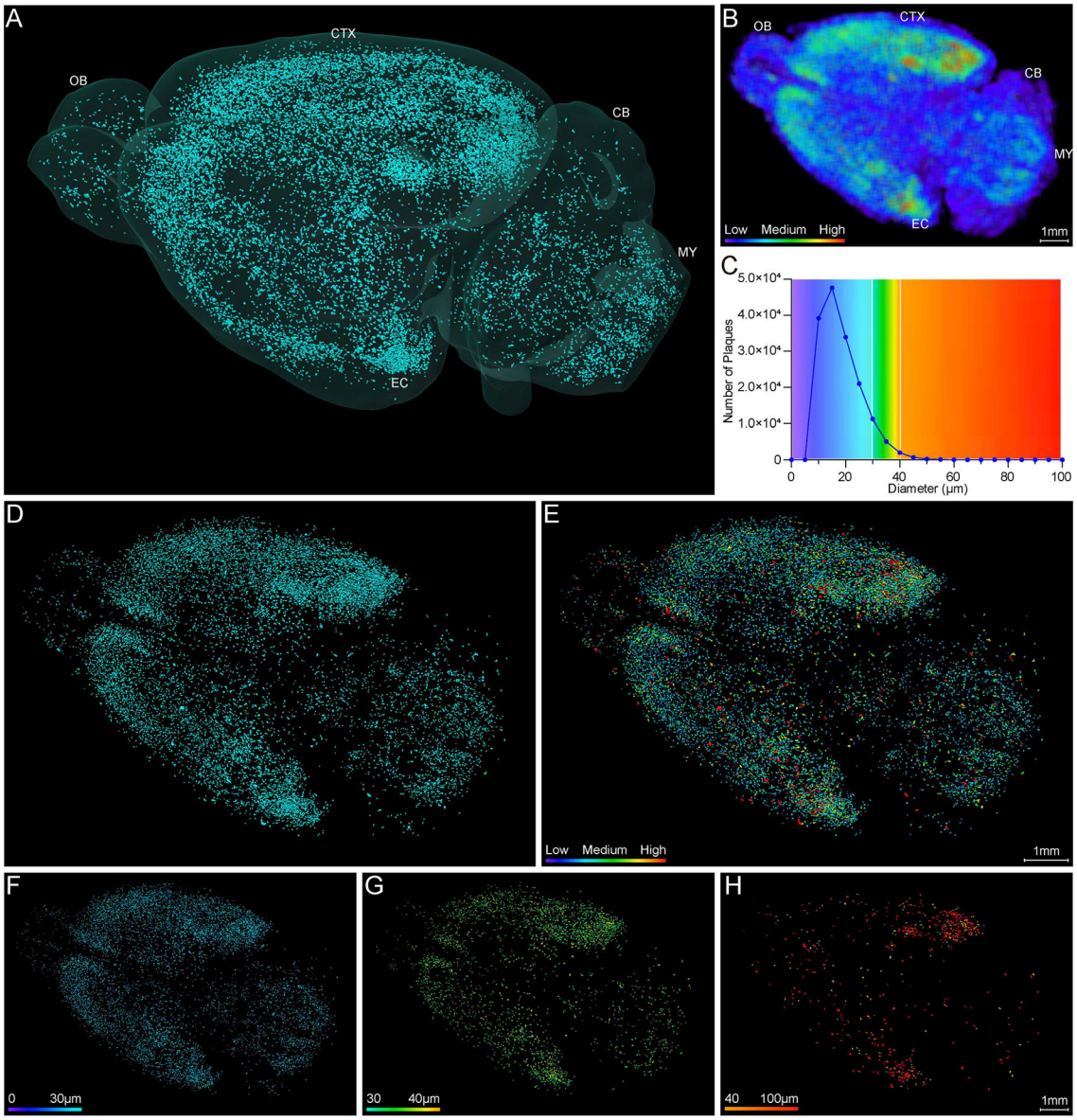


FIGURE 3 | Whole-brain 3D visualization of anatomical distribution of Aβ plaques. **(A)** 3D lateral view of the whole brain amyloid plaque distribution in the 5×FAD mice. **(B)** A 3D heatmap of plaque density distribution from apical view. The density of plaques was coded by a color scale and deep blue represents the lowest density. **(C)** Statistical analysis of the number of plaques in different diameters. The colors above x-axis were coded by the diameter of plaques, which were displayed according to different diameter ranges. **(D,E)** Top view of the amyloid plaque distribution throughout the whole brain and corresponding diameter distribution coded by a color scale and deep blue represents the lowest diameter. **(F–H)** The whole-brain plaques in different diameter ranges are separately shown: $< 30\,\mu m$ **(F)**, 30–40 μm **(G)**, $> 40\,\mu m$ **(H)**. OB, olfactory bulb; CTX, cerebral cortex; CB, cerebellum; MY, medulla; EC, entorhinal cortex.

plaques between 30 and 40 μm in diameter were relatively smaller in amount, and the plaque density varied obviously among different brain regions (**Figure 3G**). The cortical regions exhibited a denser plaque distribution, but the olfactory bulb, cerebellum, and deeper brain regions of the cerebrum had less plaques than the cortical regions (**Figure 3G**). In contrast to that most plaques were less than 40 μm in diameter, only few Aβ plaques exceeded 40 μm in diameter and were mainly distributed in the regions of the entorhinal cortex and adjacent subiculum (**Figure 3H**). The total and average plaque volumes in the regions of entorhinal cortex and adjacent subiculum were also much larger than those in the adjacent cortex and hippocampal CA3 regions (**Tables 2**, **3**). The highest density and largest size of plaques in the entorhinal cortex and adjacent subiculum suggested that Aβ plaques might first appear in these related regions and gradually increase with age.

TABLE 1 | Comparison of plaque numbers between the Entorhinal cortex and adjacent cortex regions, as well as between the hippocampal subiculum and CA3 regions.

Brain regions		Plaque number		
		Sample 1	Sample 2	Sample 3
Cortex	Entorhinal cortex	655	484	331
	Adjacent cortex	218	130	112
Hippocampus	Subiculum	826	634	573
	CA3	130	101	111

*The plaque numbers of these brain regions were calculated for three voxel blocks (0.5 × 0.5 × 0.5 mm^3, one block per mouse) from three mice. The voxel blocks were selected and cropped from the similar anatomical position of the brain of these three mice. These three voxel blocks were also employed for calculating total and average plaque volumes (**Tables 2**, **3**).*

TABLE 2 | Comparison of total plaque volumes between the Entorhinal cortex and adjacent cortex regions, as well as between the hippocampal subiculum and CA3 regions.

Brain regions		Total plaque volume (μm^3)		
		Sample 1	Sample 2	Sample 3
Cortex	Entorhinal cortex	6.76E+06	5.41E+06	1.92E+06
	Adjacent cortex	1.25E+06	4.50E+05	2.42E+05
Hippocampus	Subiculum	7.30E+06	4.71E+06	6.75E+06
	CA3	2.43E+06	1.97E+05	2.60E+05

TABLE 3 | Comparison of average plaque volumes between the Entorhinal cortex and adjacent cortex regions, as well as between the hippocampal subiculum and CA3 regions.

Brain regions		Average plaque volume (μm^3)		
		Sample 1	Sample 2	Sample 3
Cortex	Entorhinal cortex	1.03E+04	1.12E+04	5.80E+03
	Adjacent Cortex	5.73E+03	3.46E+03	2.16E+03
Hippocampus	Subiculum	8.84E+03	7.43E+03	1.18E+04
	CA3	1.87E+04	1.95E+03	2.34E+03

Synchronous Visualization of Aβ Plaques, Somata, Nerve Tracts, and Blood Vessels in Whole Mouse Brain

To ascertain the spatial correlation between Aβ plaques and their surrounding brain structures, we visualized the Aβ plaques, somata, nerve tracts, and blood vessels simultaneously in the intact brain of 5×FAD mice. **Figure 4A** shows the colocalization of somata (green), blood vessels (red), nerve tracts (purple), and Aβ plaques (cyan) across the entire brain. The high-precision rendering of whole brain revealed the 3D images of multiple brain components with structure-specific pseudo-colors. Owing to the presence of massive dense granule cells, the visualization of somata (green) showed distinctive cytoarchitectonic features in the olfactory bulb and cerebellum (**Figures 4A–C**). **Figure 4B** shows the spatial distribution difference among Aβ plaques, somata, and nerve tracts. Apparently, the nerve tracts principally spread over the WM areas, whereas the somata were mainly distributed in the gray matter (GM) areas. In the cerebrum, the Aβ plaques were densely distributed in the deep layer other than the superficial layer of the cortex, and the distribution of Aβ plaques was correlated with that of somata and nerve tracts (**Figures 4B–E**). The Aβ depositions were obviously much denser in the GM areas where were full of somata, as compared to those in the WM areas where full of tracts (**Figures 4D,E**). In addition, the spatial distribution of Aβ plaque also showed a correlation with the vascular distribution. Plaque-dense areas were generally localized in the distal end rather than root of vessels (**Figures 4C,F**).

The representative images of coronal slices across the striatum (**Figures 4G–I**) and hippocampus (**Figures 4J–L**) also exhibited the similar Aβ plaque distribution pattern. From these local high-resolution perspectives, the plaques were relatively enriched in soma-dense areas (**Figures 4G,J**) and were sparse in tract-dense areas (**Figures 4H,K**). The blood vessels were largely arranged in parallel and penetrated the deep cortical layer. The plaques were mainly aggregated in the distal end of the blood vessels (**Figures 4I,L**). In conclusion, the plaque-dense areas exhibited more somata but less nerve tracts and localized in the distal end of vessels.

Spatial Distribution Correlation of Aβ Plaques, Somata, and Blood Vessels in Cortex and Hippocampus

Aβ Plaque, Somata, and Blood Vessels in Cortex

To further investigate the correlation of Aβ plaques, somata, and blood vessels in local regions, we performed a 3D volume rendering of the cortical voxel blocks. The voxel block cropped from the cortical area as shown in **Figure 5A** exhibited the representative cerebral architecture in the 5×FAD mice. **Figures 5B–I** depict the images of the same section as shown in **Figure 5A** over a wider range of gray levels (color-coded) that revealed Aβ plaques, somata, and blood vessels. **Figures 5B,C** revealed 3D images of Aβ plaques, somata, and blood vessels with structure-specific pseudo-colors. By viewing somata and plaques (**Figure 5D**), we confirmed that the Aβ plaques were prone to accumulate in the deeper layer of the cortex, which could also be observed from **Figures 3B**, **4B,D,G,J**. The deeper layer of the cortex had more abundant somata and distal vessels (**Figures 5D–I**). The visualization of somata and vasculatures in WT mice also depicted the similar distribution pattern (**Supplementary Figures 2A–E**). The normal cortex showed an apparently layered distribution of somata, and the soma number varied in different layers, reaching the highest in the deeper layer (**Supplementary Figures 2C,E**). In addition, we performed a quantitative analysis on the number of somata and plaques with cortical depth changing based on the images of 200-μm-thick coronal sections similar to that shown in **Figure 5J**. A total of three 5×FAD mice samples were employed to calculate the spatial relationship between somata and plaques, and the x-coordinate of cortical depth was normalized by the whole cortical size (**Figure 5K**). The quantitative results revealed that

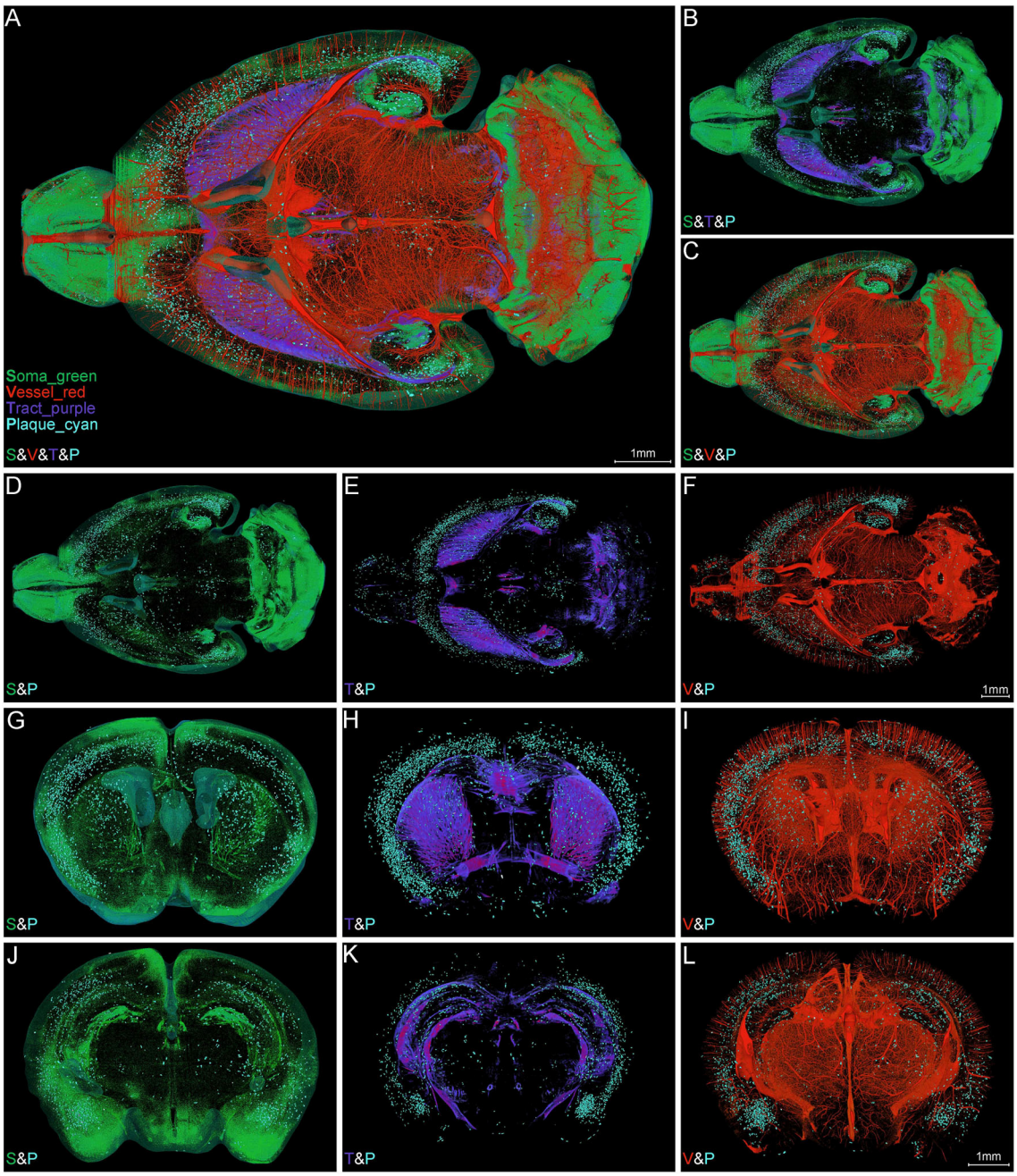


FIGURE 4 | Simultaneous whole-brain visualization of Aβ plaques, somata, nerve tracts, and blood vessels. **(A–F)** Simultaneous whole-brain visualization of multiple signals in the 5×FAD mice with different structures coded by different pseudo-colors. The "S," "V," "T," or "P" in the left bottom of each image refer to soma, vessel, tract and plaque, respectively (the same below). **(A)** Soma, vessel, tract, and plaque. **(B)** Soma, tract, and plaque. **(C)** Soma, vessel, and plaque. **(D)** Soma and plaque. **(E)** Tract and plaque. **(F)** Vessel and plaque. **(G–L)** Simultaneous visualization of multiple signals in the representative coronal slices across striatum **(G–I)** and hippocampus **(J–L)** in the same mouse as shown in **(A)**. **(G,J)** Soma and plaque. **(H,K)** Tract and plaque. **(I,L)** Vessel and plaque. Soma, vessel, tract, and plaque are shown in green, red, purple, and cyan, respectively (the same below).

the depth with the largest number of plaques was in the vicinity of that with the largest number of somata, which demonstrated that the plaques were prone to distribute near the location that more somata gathered. These results demonstrated that there was a strong spatial distribution correlation between Aβ plaques and somata.

Aβ Plaque, Somata, and Blood Vessels in Hippocampus

The representative images of hippocampal regions also depicted a characteristic Aβ plaque distribution. **Figures 6A–I** show the Aβ plaques, somata, and blood vessels in 200-μm-thick coronal sections across the hippocampus. The plaques

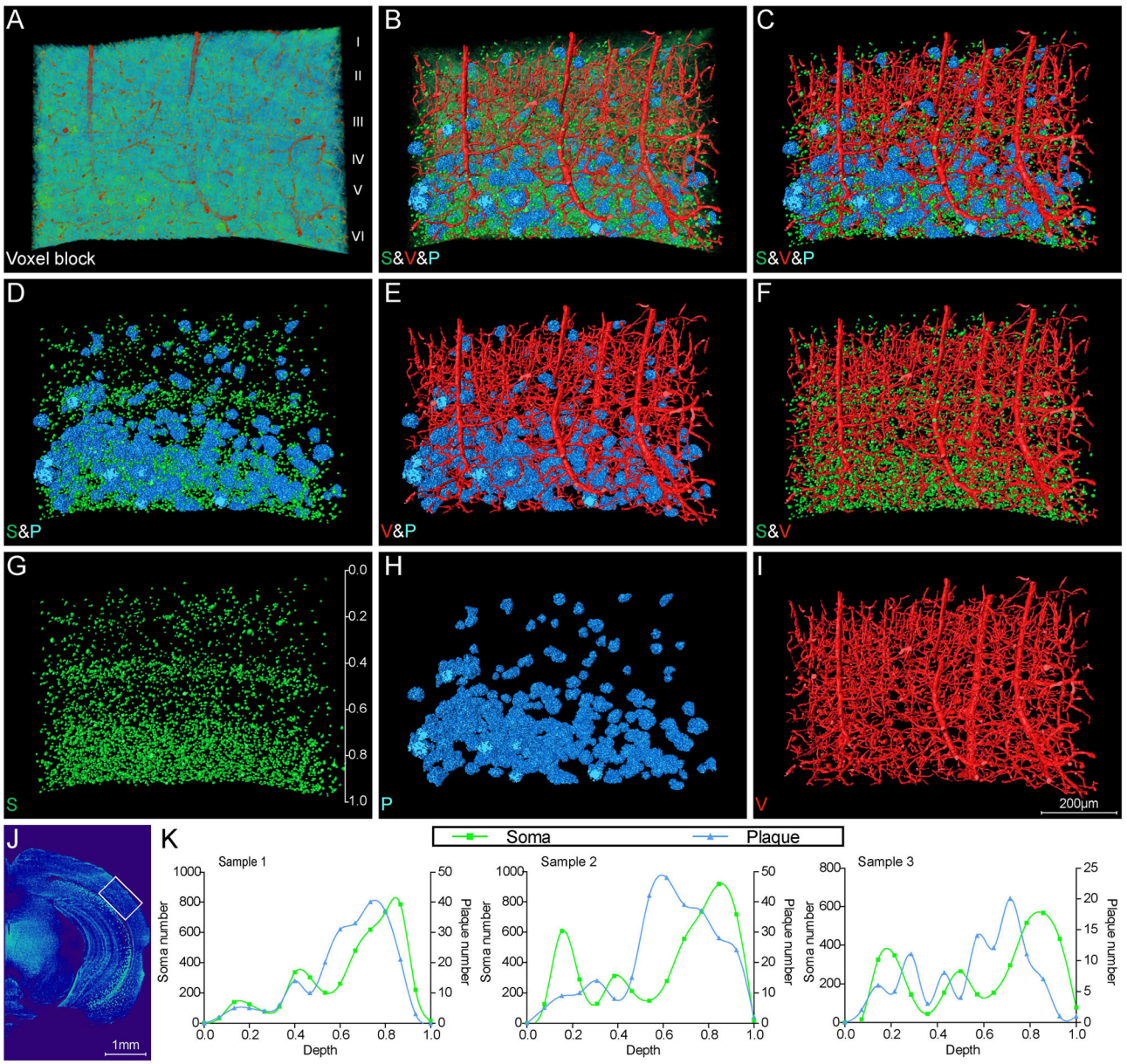


FIGURE 5 | Visual reconstruction of Aβ plaques and adjacent somata and vessels in cortical area. **(A)** Representative voxel block from the cortical region with default colormaps in the 5×FAD mice. **(B–I)** The same voxel blocks as shown in A were displayed with plaques, somata, and vessels to illustrate the spatial relationship among these signals. **(J)** The boxed area indicated the position of the voxel block as shown in **(A–I)**. **(K)** Quantitative analysis of the number of somata and plaques with cortical depth changing based on images of 200-μm-thick coronal slices similar to that shown in **(J)**. The left y-axis denoted the number of somata and the right y-axis denoted the number of plaques. The x-axis represented cortical depth which was normalized by the whole cortical size of each sample.

were predominantly distributed in the stratum oriens of cornu ammonis and the molecular layer of dentate gyrus (**Figures 6A–C**). It should be noted that the plaques in the stratum oriens and molecular layer were in close proximity to the stratum pyramidale and granule cell layer, as well as located in the distal end of blood vessels. These results indicated the distribution correlation of the plaques, somata, and blood vessels in the hippocampus, which was also in accordance with the visualization results in the whole brain (**Figures 4B–D,F**). There were few plaques appearing in the stratum radiatum, which was full of neuronal processes (**Figure 6A**). High-precision volume rendering of 400-μm-thick slices similar to the same coronal slice as shown in **Figure 6A** also described the characteristic plaque distribution (**Figures 6D–F**). By visualizing the plaques and vessels, we found that more abundant plaques were distributed in the distal end rather than in the root of vessels (**Figure 6D**). The normal distribution relationship of hippocampal structures is shown in **Supplementary Figures 2F–J**, which demonstrated the

specific distribution of somata, vessels, and neuronal processes in different hippocampal subregions. To further define the spatial relationship between plaques and somata, we extracted the plaques and their neighboring cell layers including both pyramidal cell and granule cell layers (**Figures 6E–I**). The volume rendering of hippocampal voxel block in **Figures 6E,F** showed the stratiform plaques in the vicinity of cell layers. From the extracted images, we found some plaques embedded in the cell layer, which leads to a notch in the intact cell layer (**Figures 6G–I**). In conclusion, these results revealed a characteristic layered distribution pattern of Aβ plaques in both cortical and hippocampal regions.

Structural Alternations of Nerve Processes and Capillaries Surrounding Aβ Plaques at Subcellular Resolution

Abnormalities of Neuronal Processes Surrounding Aβ Plaques

To draw a precise map of the plaque deposition and adjacent nerve processes, we made simultaneous visualization of Aβ plaques and nerve processes at subcellular resolution in the local hippocampal regions. **Figures 7A–D** show the representative images of coronal and longitudinal sections of nerve processes in the right hippocampus of the WT mice. The processes in the stratum radiatum appeared compact spoke-like extensions deriving from the somata in the stratum pyramidale (**Figures 7B,D**). Compared with those processes in the WT mice, the processes showed apparent abnormalities in the 5×FAD mice (**Figures 7E–H**). The plaques were dispersed over the stratum oriens, stratum radiatum, and subiculum. It is noted that the processes were nearly absent in areas where plaques aggregated heavily. The processes were remarkably reduced in the subiculum region (**Figures 7G,H**), which was also consistent with the highest plaque density in the subiculum region as shown in **Figure 3B**. To further focus on typical individual Aβ plaque and surrounding nerve processes, we performed a visual reconstruction of typical region in the stratum radiatum of cornu ammonis at subcellular resolution (**Figures 7I–K**). The processes passing through or near Aβ plaques exhibited various abnormalities including abrupt branch ending or bending. In addition, we found that a cluster of smaller somata, generally identified as microglia cell or astrocytes in the previous imaging studies (Liu et al., 2010; Serrano-Pozo et al., 2011), was dispersed in or around the Aβ plaques (**Supplementary Figures 3A–F, S4A–I**). These plaque-associated somata were smaller in size and appeared darker in grayscale Nissl staining images (**Supplementary Figures 3B–F**).

Deformation of Capillaries Associated With Aβ Plaques

To clearly analyze the interaction between Aβ plaques and vasculatures in local regions, we visualized the plaques and vasculatures simultaneously in the hippocampus. The distribution of plaques and vasculatures across the entire unilateral hippocampus is shown in **Figure 8A**. The plaques were found throughout the hippocampus. Besides, the plaque-denser areas in the ventral part exhibited a reduction of blood vessels, and the plaque-sparser areas in the dorsal part showed relatively richer blood vessels (**Figure 8A**). These results demonstrated the substantial disruption of the hippocampal angioarchitecture. Actually, with respect to the cortical architecture, the plaque led to the holes in the compact network of nerve cells (**Supplementary Figures 4A–C**), and the capillaries were impaired heavily with severe fragmentation and truncation (**Supplementary Figures 4D–F**). To further present the detailed fine-scale description of the vasculature, a representative view of cross-sections perpendicular to the longitudinal hippocampal axis is displayed (**Figures 8B–D**). The Aβ plaques aggregated heavily in the region of the subiculum (**Figures 8B,C**), in which the vasculature was accordingly damaged severely (**Figures 8B,D**). The cross-view of normal vasculature in WT mice is shown in **Supplementary Figure 2I**, with no plaques in the vascular network. The magnification of the subiculum region also depicted the spatial relationship between plaques and blood vessels (**Figures 8E–G** and **Supplementary Movie 3**). The plaques in different sizes were hanged over the branches of dense capillary network. From high-resolution images of the subiculum (**Figures 8H–J**), we found that capillaries inside or adjacent to the plaques appeared with apparent deformations including distorted micro-vessels and abrupt ending. In addition, the plaques in the compact network of capillaries and nerve cells exhibited more holes and bulges, which lead to the excessive irregularities in surface and high level in roughness (**Figure 8I** and **Supplementary Figures 4G–I**). It should be noted that the plaques in the dense network of processes located far from the dense cell layers and appeared relatively solider and smoother, which is shown in **Figures 7I–K**.

DISCUSSION AND CONCLUSION

In this study, we constructed an all-in-one panorama of Aβ plaques, somata, nerve processes and tracts, and blood vessels in the whole mouse brain of AD mice with a non-fluorescent labeling method. To achieve the simultaneous visualization of Aβ plaques and their surrounding brain structures, a customized image processing workflow coupled with the MOST bright-field imaging and whole-brain Nissl staining method were developed and carried out. With this technique route, we are capable to discern characteristic signals of Aβ plaques, somata, nerve processes and tracts, and blood vessels based on the difference of gray values and morphologies in the whole-brain Nissl staining dataset.

Identification of Whole-Brain Multiple Structures Without Specific Labeling

High-content fluorescence imaging highly relies on the acquisition of multi-channel signals simultaneously. *In vivo* fluorescence labeling methods, which include injections of fluorescent dyes (Robertson et al., 2015), injection with virus

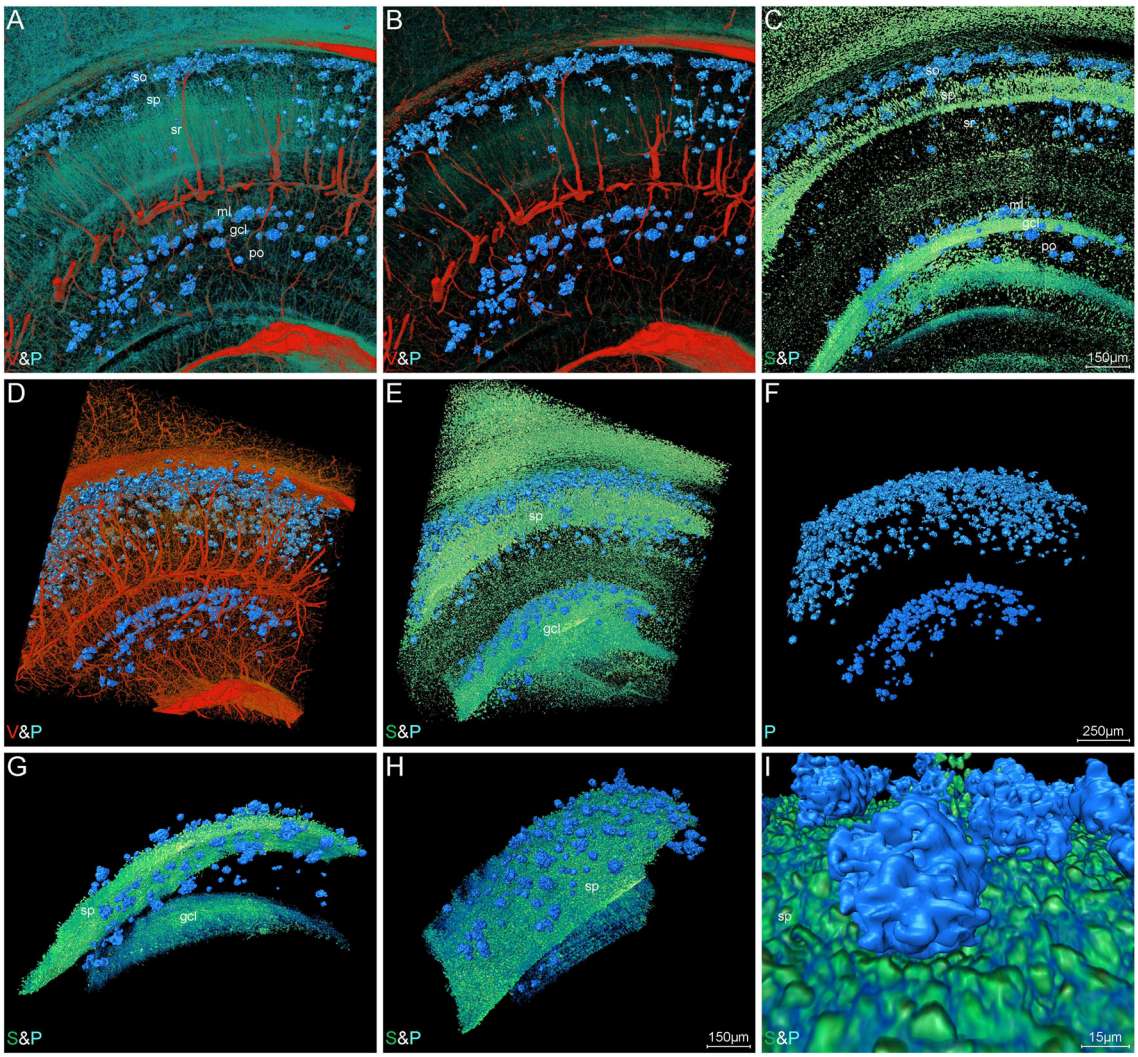


FIGURE 6 | Visual reconstruction of Aβ plaques and adjacent somata and vessels in hippocampal area. **(A–C)** Representative coronal view of 200-μm-thick slices showing Aβ plaques and other structures in the hippocampus of 5×FAD mice. **(D–I)** 3D volume rendering of 400-μm-thick sections across the same coronal slice as shown in A describing the specific plaque distribution. **(D)** The Aβ plaques were dispersed abundantly in the areas of the distal end of blood vessels. **(E,F)** The Aβ plaques were prominently distributed into a layer and the majority were close to the cell layer, namely stratum pyramidale and granule cell layer. **(G,H)** Lateral view of the extracted cell layer, with stratiform Aβ plaques visible adjacently to the cell layer. **(I)** The magnified view showing the Aβ plaques embedded in the cell layer. In the Cornu Ammonis, so, stratum oriens; sp, stratum pyramidale; sr, stratum radiatum. In the dentate gyrus, ml, molecular layer; po, polymorph layer; gcl, granule cell layer.

expressing fluorescent proteins (Zingg et al., 2017), and transgenic mice expressing fluorescent proteins (Wessels et al., 2008), are capable of labeling several different brain structures simultaneously. For instance, Prox1-GFP/Flt1-DsRed mice with inherent fluorescence of blood and lymphatic vessels allow for direct visualization of blood and lymphatic vessels in various organs (Zhong et al., 2017). Combining with fast-developing brain-wide imaging techniques, such as light sheet microscopy (Dodt et al., 2007), serial two-photon tomography (Ragan et al., 2012), and fMOST (Gong et al., 2016), the whole-brain imaging of specimen having two or three fluorescent labels is achievable. However, imaging of specimens with three or more fluorescent labels is often complicated by the bleed-through or crossover of fluorescence emission, unless the spectral profiles of the fluorophores are very well separated (Valm et al., 2016; McRae et al., 2019).

Here, we utilized the whole-brain Nissl staining, a non-fluorescent labeling method, to achieve simultaneous displaying of whole-brain multiple structures in one sample. The conventional Nissl staining was a classic nucleic acid staining method directing at thin tissue sections and was commonly adopted to label somata with displaying Nissl body or revealing reduction in neuronal population (Nobakht et al., 2011), whereas the whole-brain Nissl staining method was modified to better label an entire brain. After the whole-brain

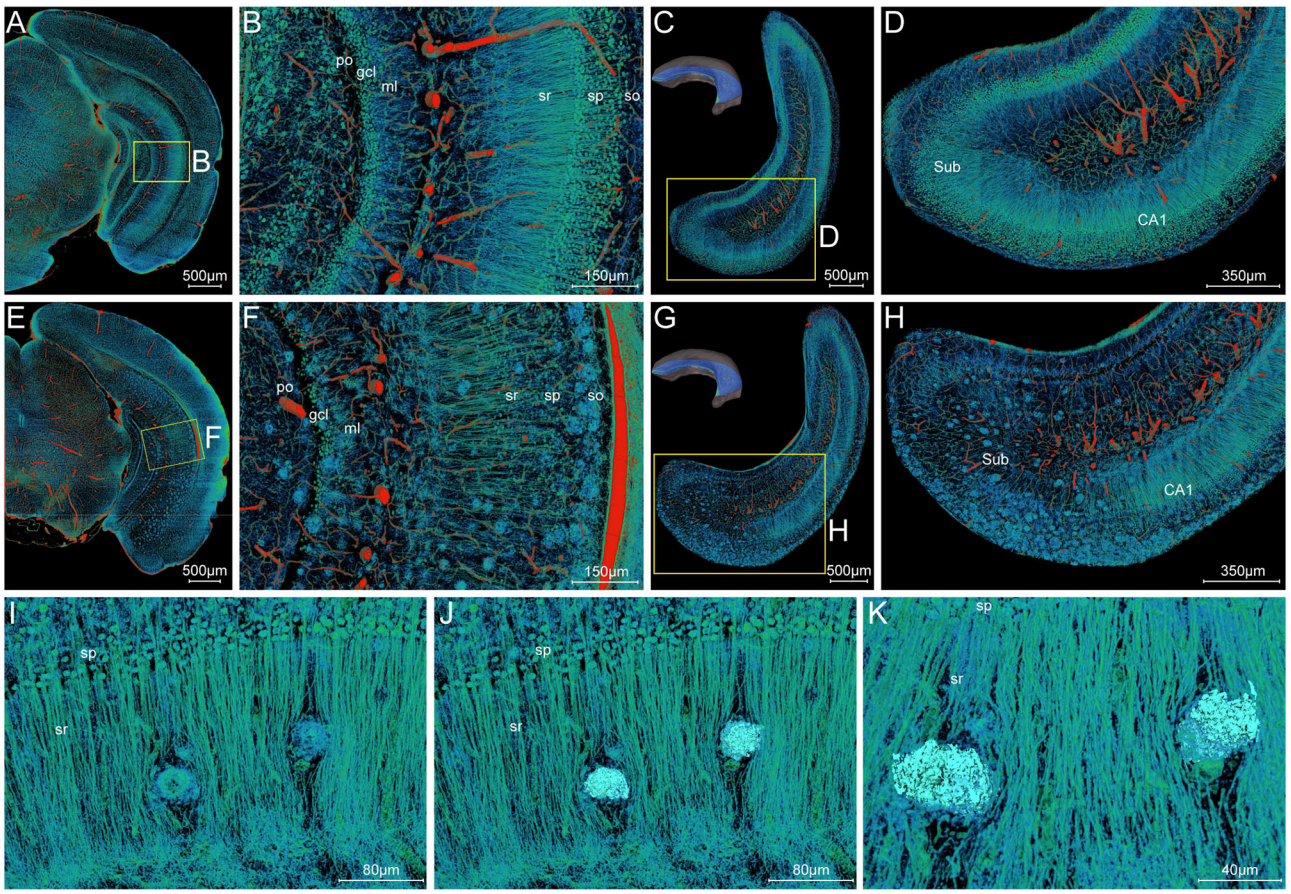


FIGURE 7 | Visualization of neuronal processes surrounding Aβ plaques. **(A–D)** Simultaneous visualization of neuronal processes and somata in the WT mouse. **(E–H)** Simultaneous visualization of Aβ plaques, neuronal processes, and somata in the 5×FAD mouse. **(A,B,E,F)** Representative image of the typical coronal slice across the hippocampus and the magnification of the boxed area. **(C,D,G,H)** Representative image of the typical slice parallel to the longitudinal hippocampal axis across the hippocampus and the magnification of the boxed area. Sub, subiculum. **(C,G)** The unilateral hippocampus in the upper left corner indicated the location of the reconstructed slices parallel to the longitudinal hippocampal axis. **(I–K)** Representative high-resolution images of the strata radiatum in the hippocampus showing the typical Aβ plaques and their surrounding neuronal processes. The same view showed the Aβ plaques displayed by volume rendering **(I)** and surface reconstruction **(J)** in the compact network of neuronal processes. The more clear view in **(K)** showed the neuronal processes with abrupt branch ending or bending. In the Cornu Ammonis, so, stratum oriens; sp, stratum pyramidale; sr, stratum radiatum. In the dentate gyrus, ml, molecular layer; po, polymorph layer; gcl, granule cell layer.

Nissl staining, the Aβ plaques, somata, nerve processes and tracts, and blood vessels would show varying degrees of gray values and could be identified and extracted integrating with their different morphologies. Noticeably, the application of bright-field Nissl staining to display Aβ plaque has not been verified in the other studies. Our study achieved the first whole-brain distribution of Aβ plaques based on the Nissl staining. We speculated that several factors contributed to the clear presentation of Aβ plaque in this study. First, the ultrathin sections (1 μm) employed in the MOST system resulted in high axial resolution and the modification of whole-brain Nissl staining method led to enhanced contrast, both of which made it more feasible to display the Aβ plaques. Importantly, we also adjusted the MOST system with relative higher image contrast and kept the image sequence being high-quality during the long-term imaging. Besides, the capability of bright-field image processing enabled us to make 3D reconstruction on the pompon-like signal of interest, which may also contribute to the recognition and verification of the Aβ plaque in Nissl staining. In the conventional Nissl staining brain slices, the redundant information except the somata was usually ignored, even though the Aβ plaque might be displayed occasionally. Our study suggested several benefits of the application of whole-brain Nissl staining and MOST acquisition. First, analyzing the Aβ plaques in bright field images yields the possibility to measure more descriptive features such as texture and shape simultaneously. Additionally, we thus found that the Aβ plaques surrounding by nerve processes and few somata were solider and smoother, but those in compact network of capillaries and nerve cells appeared more irregular and rougher. Second, the simultaneous acquisition of multiple signals also breaks the limitation of spectral overlap of the fluorophore. To sum up, the whole-brain Nissl staining provides an alternative non-fluorescent method for labeling multiple components in

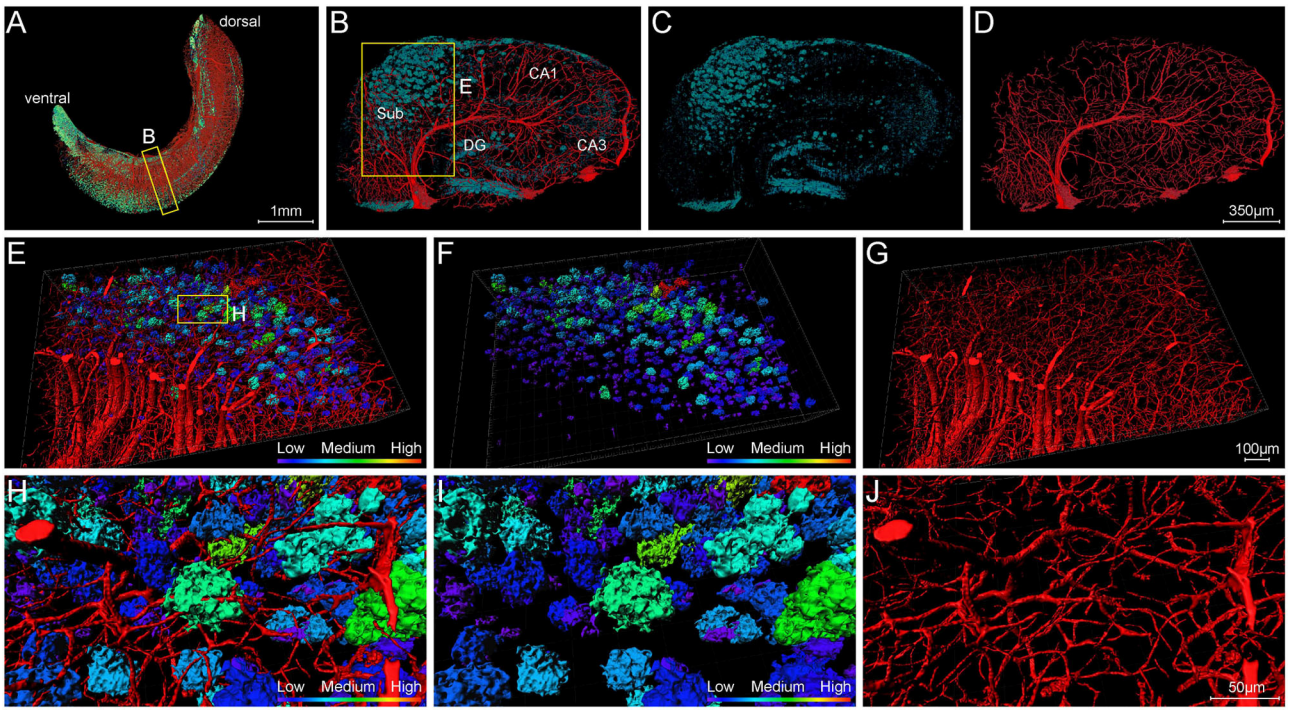


FIGURE 8 | Visualization of blood vessels surrounding Aβ plaques. **(A)** Front view of Aβ plaques and blood vessels in the right hippocampus showing the abundant plaques in the ventral part of the hippocampus of 5×FAD mice, with vessels eliminating. The sectional view of the box area in **(A)** is enlarged and shown in **(B)**. **(B–D)** Representative sectional views of hippocampal vasculature and plaque burden perpendicular to the longitudinal hippocampal axis. **(B)** Vessel and plaque. **(C)** Plaque. **(D)** Vessel. **(E–G)** The magnifications of the box in B showing the plaques in the compact vascular network. **(E)** Vessel and plaque. **(F)** Plaque. **(G)** Vessel. **(H–J)** The magnifications of the box in E depicting the interaction of plaques and vessels. **(H)** Vessel and plaque. **(I)** Plaque. **(J)** Vessel.

brain research and offers new idea for the conventional method in new use.

Customized Image Processing Workflow for Simultaneous Visualization

The MOST system adopted the reflected bright field imaging to generate a massive amount of data (Li et al., 2010). The processing method for the high-throughput bright-field images is more demanding than in the fluorescence case (Buggenthin et al., 2013). The bright-field images exhibited more heterogeneous gray levels and lower contrast than the fluorescent images. Specifically, the gray value of Aβ plaques exhibited inhomogeneous distribution of "W" shape and was close to that of the brain parenchyma. This characteristic gray-value distribution of Aβ plaque made it difficult to be extracted. Besides, the inhomogeneous gray level may also exist among slices over the long periods of imaging. Therefore, despite the multiple signals of different structures were provided by the whole-brain Nissl staining dataset generated from the MOST system, there are still considerable obstacles faced by simultaneous visualization of these multiple structures using direct volume rendering.

In this study, we have made attempts in developing a customized image processing workflow for the extraction and reconstruction of these multiple structures. An image optimization method for high-throughput bright-field images was first performed to enhance the image contrast and to facilitate the following segmentation. To simultaneously visualizing all of these signals in the same whole brain, we took advantage of the distinct features of different structures and utilized virtual channel splitting to extract the information of different structures, as well as ultimately performed feature fusion to provide a visual reconstruction of multiple structures. This visualization workflow could be adopted in other high-throughput bright field imaging data and will also greatly benefit the systematic understanding of the complex brain. The ability to visualize multiple structures at submicron resolution of 3D whole-brain scale, especially in the same sample, will help to clarify the underlying mechanisms of neurodegeneration.

All-in-One Panorama of Multiple Structures in Whole Brain of AD Mice

Aβ plaques, currently believed as the earliest detectable AD biomarker, are increasingly recognized to be an intricated role interplaying with various neighboring brain structures such as neurons, glia cells, and vasculatures (Marin et al., 2016; Yuan et al., 2016; Miners et al., 2018; Greenberg et al., 2020). The previous studies employing brain-wide imaging techniques have realized the brain-wide visualization of Aβ plaques (Liebmann et al., 2016; Long et al., 2019; Whitesell et al., 2019). Besides,

whole-brain imaging of the single surrounding structure such as blood vessels, cells, and neural circuits associated with Aβ plaques was also achieved easily (Meyer et al., 2008; Liebmann et al., 2016; Zhang et al., 2020). Nevertheless, a few studies have systematically correlated the other multiple surrounding structures morphology with their spatial relation to Aβ plaques, especially at the submicron level in the intact brain. Our study provided a novel and robust approach for simultaneous cross-scale reconstruction of Aβ plaques and their surrounding multiple brain structures in the whole brain. On the brain-wide scale, the spatial distribution of Aβ plaques was correlated with that of somata, blood vessels, and nerve tracts. In addition, the entorhinal cortex and adjacent subiculum regions that present the highest density of plaques and also large-sized plaques suggested that Aβ plaques might first appear in these related regions and gradually increase with age, which was consistent with the report that the 5×FAD mice developed rapid and severe amyloid pathology, with extracellular Aβ deposition beginning at 2 months in layer V of the cortex and subiculum (Oakley et al., 2006). In the local regions of cortex and hippocampus, the Aβ plaques were found distributed in the vicinity of cell layers and located in the distal end of blood vessels. Given that the 5×FAD mice were generated by the neuron-specific mouse Thy1 promoter to drive APP and PS1 transgenes overexpression (Oakley et al., 2006), it is not difficult to understand the neuronal origin for the Aβ deposits and the strong spatial distribution correlation between Aβ plaques and somata. By viewing at subcellular resolution, the bending or abrupt ending of the nerve processes, as well as distortion or fracture of micro-vessels, was visualized.

Our study has revealed the heterogeneous distribution of Aβ plaque even in the same brain regions (**Figures 5H,K**) and the strong correlation between Aβ plaque and soma (**Figures 5**, **6**), which implied that investigating the plaque distribution in the context of its surrounding multi-structures might be more beneficial compared with that in different brain regions. Although further study to elucidate the underlying physiological function and mechanism behind significant structural alternation in AD is needed to carried out, our findings provided the cross-scale high-precision panorama of multiple brain structures associated with AD pathology in one same sample. Our study placed emphasis on solving problems in processing high-throughput bright field images and made attempt in developing a method for the extraction and reconstruction of multiple structures. This will facilitate a better understanding of the cerebral anatomical features under the pathological state of AD and shows extensive application prospect in drug efficacy assessment from brain-wide level.

METHODS

Sample Preparation and Data Acquisition

The 6-month-old 5×FAD mice and WT littermates were provided and used in this work. The 5×FAD mice carry three mutations of the human amyloid precursor gene (Swedish-K670N/M671L, Florida-I716V, and London-V717I) and two mutations of the human presenilin-1 gene (M146L and L286V) and recapitulate many AD-related phenotypes and have a relatively early and aggressive presentation (Oakley et al., 2006). The transgenic mice were obtained from The Jackson Laboratory (stock no: 34840-JAX), and the corresponding WT control mice were littermates of the AD mice. All the animal procedures were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals, under the protocols approved by and strictly following the guidelines of the Institutional Animal Care and Use Committee (IACUC). The whole-brain staining method followed the previous description (Wu et al., 2014). In brief, the entire intact brains were carefully removed and post-fixed after cardiac perfusion. Then, the brains were stained with 2.5% thionine (Sigma-Aldrich, 861340) solution for at least 12 days. The stained brains were subsequently dehydrated with a graded series of ethanol and acetone solutions. Next, the dehydrated brains were immersed and then embedded in Spurr resin (Sigma-Aldrich, EM0300) to be suitable for the MOST system. During the long-term imaging, we adjusted the MOST system with high image contrast and kept the image sequence being high quality.

Image Preprocessing and Optimization

The raw images were preprocessed for seamless image stitching and luminance nonuniformity correction based on a previous preprocessing program (Ding et al., 2013). Then, an image optimization method was performed to further improve background uniformity, periodic noise, and to enhance the image contrast of the preprocessed coronal images. The optimization was implemented through steps including background correction, noise reduction, and contrast enhancement (Zhang et al., 2019). The enhanced contrast of the optimized coronal images significantly improved the convenience of information extraction and grayscale rendering.

Verification of Aβ Plaque

The brains were first prepared by the whole-brain Nissl staining method without resin embedding and then followed by Aβ-specific staining of paraffin sections. Briefly, the mouse brains were first put into 2.5% thionine staining solution for 12-day whole-brain Nissl staining. After the staining completed, the mouse brains were immersed in 50% ethanol for 2 h and then in 70% ethanol for 6 days. The samples were subsequently subjected to gradient dehydration and paraffin infiltration: 95% ethanol for 2 h; two steps of 100% ethanol for 1 h each; mixed volume of xylene and absolute ethanol for 20 min; two steps of xylene for 20 min each; dipping wax (56–58°C) for 2 h; embedding; slicing. After obtaining the paraffin sections of 5×FAD mouse brains, anti-Aβ antibody ab2454 (CST, 2454s) and Alexa fluor 546-conjugated secondary antibody were incubated on the same sections. Subsequently, the sections were stained with 1% ThS solution and mounted with a coverslip. Finally, the sections were imaged with confocal microscopy.

Simultaneous Visualization of Multiple Brain Components

The optimized high-quality coronal images of 5×FAD mice were directly used to extract different brain structure information in different gray levels. The enhanced contrast of the optimized

images significantly improves the convenience of recognition and extraction of Aβ plaques, blood vessels, somata, neuronal processes, and nerve tracts. In this paper, the structural information was classified into three categories, namely, blood vessels, plaques, and neural structures, which were extracted by virtual channel splitting individually (refer to **Supplementary Figure 1**). The source code is available and provided in the **Supplementary Material**. The code includes scripting commands using Tool Command Language (Tcl) with Amira-specific extensions. It allows users to automate certain processes and to create scripts for managing routine tasks or for presenting demos in the Amira software (Amira 2019.2). In brief, before executing the Tcl code, the data directory path must be set in the scripts. The location where should input the data directory path had been annotated in the word file "Scripts for Virtual channel splitting." Then, copy all the scripts to a TXT file and change the file extension as ".hx." Next, drag and drop the ".hx" file onto the Amira console to run the Tcl code in the Amira software. It should be noted that the parameters set in the scripts are suitable for the dataset generated by MOST and optimized by our method (Zhang et al., 2019). When handling with the other datasets, the parameters should be correspondingly modified.

Subsequently, high-quality rendering of these different structures was integrated through feature fusion to accomplish synchronous visualization of multiple structures in one sample. Iso-surface rendering and direct volume rendering were employed to achieve high-quality rendering of brain structures using Amira software. Iso-surface rendering intends to visualize an iso-surface in the 3D data field (Bosma et al., 1998) and has several well-recognized advantages in efficiency and accuracy. Direct volume rendering represents one of the most widely used methods for 3D visualization of tomographic images. In volume rendering, image projections are constructed by simulating the absorption and emission of light in an image stack along each ray path to the eye. The blood vessels were filtered out by region growth algorithm and then were visualized using iso-surface rendering. The somata, nerve processes, and tracts were reconstructed using direct volume rendering in Amira software. The pompon-like structures were visualized by high-precision real-time rendering after feature enhancement.

We first reconstructed Aβ plaques, blood vessels, somata, and nerve tracts at the whole brain scale and then reconstructed these various brain structures on the typical coronal plane (400 μm in thickness) of the striatum and hippocampus. The high-precision reconstruction of somata, blood vessels, and Aβ plaques in the typical cortical and hippocampal regions was also conducted. The impact of Aβ plaque on its surrounding nerve processes and capillaries was investigated by visualizing nerve processes or blood vessels associated with Aβ plaques. A 3D heatmap of plaque density was performed across the whole brain. The density of plaques was coded by a color scale and deep blue represents the lowest density. The plaque number and volume in the entorhinal cortex and its adjacent cortex, as well as in the hippocampal subiculum and CA3 regions, were calculated for three voxel blocks ($0.5 \times 0.5 \times 0.5$ mm^3, one blocks per mouse) from the similar anatomical position of the brain of three mice. Besides, the soma number and plaque number for three 5×FAD mice were calculated in different cortical depths that were normalized by the whole cortical size of each sample. The soma number for WT mice was also calculated in different cortical depths to evaluate the normal distribution correlation between somata and blood vessels.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s. The source code is provided in the **Supplementary Material** (refer to Scripts for Virtual channel splitting).

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of Shanghai Institute of Materia Medica.

AUTHOR CONTRIBUTIONS

HJ, ZG, and HZ conceived and designed the research. JZ performed the animal experiments. XZ performed the sample preparation, data acquisition, and image preprocessing. XZ and JZ performed the validation of Aβ plaque. XY performed the image optimization, visualization, and quantitative analysis. HJ, ZG, HZ, XY, XZ, and JZ wrote the manuscript. All authors analyzed the data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Science Fund for Distinguished Young Scholars (81825021 to ZG), Youth Innovation Promotion Association CAS (2018323 to XY), Shanghai Municipal Science and Technology Major Project (2018SHZDZX05 to HJ and ZG), China Postdoctoral Science Foundation (2021M703344 to XZ), and Shanghai Super Postdoctoral Incentive Program (2021 431 to XZ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.870520/full#supplementary-material

REFERENCES

Adalbert, R., Nogradi, A., Babetto, E., Janeckova, L., Walker, S. A., Kerschensteiner, M., et al. (2009). Severely dystrophic axons at amyloid plaques remain continuous and connected to viable cell bodies. *Brain* 132 (Pt 2), 402–416. doi: 10.1093/brain/awn312

Apostolova, L. G., Green, A. E., Babakchanian, S., Hwang, K. S., Chou, Y. Y., Toga, A. W., et al. (2012). Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (MCI), Alzheimer Disease. *Alzheimer Dis. Assoc. Disord.* 26, 17–27. doi: 10.1097/WAD.0b013e3182163b62

Appaix, F., Girod, S., Boisseau, S., Romer, J., Vial, J. C., Albrieux, M., et al. (2012). Specific in vivo staining of astrocytes in the whole brain after intravenous injection of sulforhodamine dyes. *PLoS ONE* 7, e35169. doi: 10.1371/journal.pone.0035169

Bacskai, B. J., Hickey, G. A., Skoch, J., Kajdasz, S. T., Wang, Y., Huang, G. F., et al. (2003). Four-dimensional multiphoton imaging of brain entry, amyloid binding, and clearance of an amyloid-beta ligand in transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12462–12467. doi: 10.1073/pnas.2034101100

Bittner, T., Burgold, S., Dorostkar, M. M., Fuhrmann, M., Wegenast-Braun, B. M., Schmidt, B., et al. (2012). Amyloid plaque formation precedes dendritic spine loss. *Acta Neuropathol.* 124, 797–807. doi: 10.1007/s00401-012-1047-8

Bolmont, T., Bouwens, A., Pache, C., Dimitrov, M., Berclaz, C., Villiger, M., et al. (2012). Label-free imaging of cerebral beta-amyloidosis with extended-focus optical coherence microscopy. *J. Neurosci.* 32, 14548–14556. doi: 10.1523/JNEUROSCI.0925-12.2012

Bosma, M., Jaap, S., and Steven, J. (1998). "Storage lobregt retrieval for image, video databases," in *Iso-Surface Volume Rendering* (San Diego, CA).

Bozzali, M., Falini, A., Franceschi, M., Cercignani, M., Zuffi, M., Scotti, G., et al. (2002). White matter damage in Alzheimer's disease assessed in vivo using diffusion tensor magnetic resonance imaging. *J. Neurol Neurosurg. Psychiatry* 72, 742–746. doi: 10.1136/jnnp.72.6.742

Buggenthin, F., Carsten, M., Michael, S., Philipp Hoppe, S., Oliver, H., Timm, S., et al. (2013). An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. *BMC Bioinformatics* 14, 297. doi: 10.1186/1471-2105-14-297

Daigle, T. L., Madisen, L., Hage, T. A., Valley, M. T., Knoblich, U., Larsen, R. S., et al. (2018). A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* 174, 465.e22–480.e22. doi: 10.1016/j.cell.2018.06.035

Delafontaine-Martel, P., Lefebvre, J., Tardif, P. L., Levy, B. I., Pouliot, P., and Lesage, F. (2018). Whole brain vascular imaging in a mouse model of Alzheimer's disease with two-photon microscopy. *J. Biomed. Opt.* 23, 1–10. doi: 10.1117/1.JBO.23.7.076501

Dickson, T. C., and Vickers, J. C. (2001). The morphological phenotype of beta-amyloid plaques and associated neuritic changes in Alzheimer's disease. *Neuroscience* 105, 99–107. doi: 10.1016/S0306-4522(01)00169-5

Ding, W., Wu, L., i,. A., Yang, J., Meng, Z., Wang, Y., et al. (2013). Automatic macroscopic density artefact removal in a Nissl-stained microscopic atlas of whole mouse brain. *J. Microsc.* 251, 168–177. doi: 10.1111/jmi.12058

Dodt, H. U., Leischner, U., Schierloh, A., Jahrling, N., Mauch, C. P., Deininger, K., et al. (2007). Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. *Nat. Methods* 4, 331–336. doi: 10.1038/nmeth1036

Dorr, A., Sled, J. G., and Kabani, N. (2007). Three-dimensional cerebral vasculature of the CBA mouse brain: a magnetic resonance imaging and micro computed tomography study. *Neuroimage* 35, 1409–1423. doi: 10.1016/j.neuroimage.2006.12.040

Enzlein, T., Cordes, J., Munteanu, B., Michno, W., Serneels, L., Strooper, D. C., et al. (2020). Computational analysis of alzheimer amyloid plaque composition in 2D-and elastically reconstructed 3D-MALDI MS images. *Anal. Chem.* 92, 14484–14493. doi: 10.1021/acs.analchem.0c02585

Fakhoury, M. (2018). Microglia and astrocytes in Alzheimer's disease: implications for therapy. *Curr. Neuropharmacol.* 16, 508–518. doi: 10.2174/1570159X15666170720095240

Fu, H., Tu, P., Zhao, L., Liu, D. A. I. J., and Cui, B. M. (2016). Amyloid-beta deposits target efficient near-infrared fluorescent probes: synthesis, *in vitro* evaluation, and *in vivo* imaging. *Anal. Chem.* 88, 1944–1950. doi: 10.1021/acs.analchem.5b04441

Gang, Y., Liu, X., Wang, X., Zhang, Q., Zhou, H., Chen, R., et al. (2017). Plastic embedding immunolabeled large-volume samples for three-dimensional high-resolution imaging. *Biomed. Opt. Express* 8, 3583–3596. doi: 10.1364/BOE.8.003583

Gao, R., Asano, S. M., Upadhyayula, S., Pisarev, I., Milkie, D. E., Liu, T. L., et al. (2019). Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. *Science* 363, 6424. doi: 10.1126/science.aau8302

Giovanna, D. I., Tibo, A. P., Silvestri, A., Mullenbroich, L., Costantini, M. C., Allegra Mascaro, I., et al. (2018). Whole-brain vasculature reconstruction at the single capillary level. *Sci. Rep.* 8, 12573. doi: 10.1038/s41598-018-30533-3

Gong, H., Xu, D., Yuan, J., Li, X., Guo, C., Peng, J., et al. (2016). High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nat. Commun.* 7, 12142. doi: 10.1038/ncomms12142

Greenberg, S. M., Bacskai, B. J., Hernandez-Guillamon, M., Pruzin, J., Sperling, R., and van Veluw, J. S. (2020). Cerebral amyloid angiopathy and Alzheimer disease - one peptide, two pathways. *Nat. Rev. Neurol.* 16, 30–42. doi: 10.1038/s41582-019-0281-2

Gurol, M. E., Becker, J. A., Fotiadis, P., Riley, G., Schwab, K., Johnson, K. A., et al. (2016). Florbetapir-PET to diagnose cerebral amyloid angiopathy: a prospective study. *Neurology* 87, 2043–2049. doi: 10.1212/WNL.0000000000003197

Ji, M., Arbel, L., Zhang, M., Freudiger, L., Hou, C. W., Lin, S. S., et al. (2018). Label-free imaging of amyloid plaques in Alzheimer's disease with stimulated Raman scattering microscopy. *Sci. Adv.* 4, eaat7715. doi: 10.1126/sciadv.aat7715

Jing, D., Zhang, S., Luo, W., Gao, X., Men, Y., Ma, C., et al. (2018). Tissue clearing of both hard and soft tissue organs with the PEGASOS method. *Cell Res.* 28, 803–818. doi: 10.1038/s41422-018-0049-z

Johansson, P. K., and Koelsch, P. (2017). Label-free imaging of amyloids using their intrinsic linear and nonlinear optical properties. *Biomed. Opt. Express* 8, 743–756. doi: 10.1364/BOE.8.000743

Kadowaki, H., Nishitoh, H., Urano, F., Sadamitsu, C., Matsuzawa, A., Takeda, K., et al. (2005). Amyloid beta induces neuronal cell death through ROS-mediated ASK1 activation. *Cell Death Differ.* 12, 19–24. doi: 10.1038/sj.cdd.4401528

Khouri, K., Xie, D. F., Crouzet, C., Bahani, A. W., Cribbs, D. H., Fisher, M. J., et al. (2021). Simple methodology to visualize whole-brain microvasculature in three dimensions. *Neurophotonics* 8, 025004. doi: 10.1117/1.NPh.8.2.025004

Kimbrough, I. F., Robel, S., Roberson, E. D., and Sontheimer, H. (2015). Vascular amyloidosis impairs the gliovascular unit in a mouse model of Alzheimer's disease. *Brain* 138 (Pt 12), 3716–3733. doi: 10.1093/brain/awv327

Kirst, C., Skriabine, S., Vieites-Prado, A., Topilko, T., Bertin, P., Gerschenfeld, G., et al. (2020). Mapping the fine-scale organization and plasticity of the brain vasculature. *Cell* 180, 780–795. e25. doi: 10.1016/j.cell.2020.01.028

Klunk, W. E., Bacskai, B. J., Mathis, C. A., Kajdasz, S. T., McLellan, M. E., Frosch, M. P., et al. (2002). Imaging Abeta plaques in living transgenic mice with multiphoton microscopy and methoxy-X04, a systemically administered Congo red derivative. *J. Neuropathol. Exp. Neurol.* 61, 797–805. doi: 10.1093/jnen/61.9.797

Konno, A., Matsumoto, N., Tomono, Y., and Okazaki, S. (2020). Pathological application of carbocyanine dye-based multicolour imaging of vasculature and associated structures. *Sci. Rep.* 10, 12613. doi: 10.1038/s41598-020-69394-0

Leiva-Salinas, C., Jiang, B., and Wintermark, M. (2018). Computed tomography, computed tomography angiography, and perfusion computed tomography evaluation of acute ischemic stroke. *Neuroimaging Clin. N. Am.* 28, 565–572. doi: 10.1016/j.nic.2018.06.002

Li, A., Gong, H., Zhang, H., Wang, B., Yan, Q., Wu, C., et al. (2010). Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain. *Science* 330, 1404–1408. doi: 10.1126/science.1191776

Liebmann, T., Renier, N., Bettayeb, K., Greengard, P., Tessier-Lavigne, M., and Flajolet, M. (2016). Three-dimensional study of Alzheimer's disease hallmarks using the iDISCO clearing method. *Cell Rep.* 16, 1138–1152. doi: 10.1016/j.celrep.2016.06.060

Liu, Z., Condello, C., Schain, A., Harb, R., and Grutzendler, J. (2010), CX3CR.1 in microglia regulates brain amyloid deposition through selective protofibrillar amyloid-beta phagocytosis. *J. Neurosci.* 30, 17091–17101. doi: 10.1523/JNEUROSCI.4403-10.2010

Long, B., Xiangning, L., Jianping, Z., Siqi, C., Wenwei, L., Qiuyuan, Z., et al. (2019). Three-dimensional quantitative analysis of amyloid plaques in the

whole brain with high voxel resolution. *Scientia Sinica Vitae* 49, 140–150. doi: 10.1360/N052019-00001

Luo, Y., Wang, A., Liu, M., Lei, T., Zhang, X., Gao, Z., et al. (2017). Label-free brainwide visualization of senile plaque using cryo-micro-optical sectioning tomography. *Opt. Lett.* 42, 4247–4250. doi: 10.1364/OL.42.004247

Mano, T., Albanese, A., Dodt, H. U., Erturk, A., Gradinaru, V., Treweek, J. B., et al. (2018). Whole-brain analysis of cells and circuits by tissue clearing and light-sheet microscopy. *J. Neurosci.* 38, 9330–9337. doi: 10.1523/JNEUROSCI.1677-18.2018

Marin, M. A., Ziburkus, J., Jankowsky, J., and Rasband, M. N. (2016). Amyloid-beta plaques disrupt axon initial segments. *Exp. Neurol.* 281, 93–98. doi: 10.1016/j.expneurol.2016.04.018

McRae, T. D., Oleksyn, D., Miller, J., and Gao, Y. R. (2019). Robust blind spectral unmixing for fluorescence microscopy using unsupervised learning. *PLoS ONE* 14, e0225410. doi: 10.1371/journal.pone.0225410

Meyer, E. P., Ulmann-Schuler, A., Staufenbiel, M., and Krucker, T. (2008). Altered morphology and 3D architecture of brain vasculature in a mouse model for Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3587–3592. doi: 10.1073/pnas.0709788105

Miners, J. S., Schulz, I., and Love, S. (2018). Differing associations between Abeta accumulation, hypoperfusion, blood-brain barrier dysfunction and loss of PDGFRB pericyte marker in the precuneus and parietal white matter in Alzheimer's disease. *J. Cereb. Blood Flow Metab.* 38, 103–115. doi: 10.1177/0271678X17690761

Miyawaki, T., Morikawa, S., Susaki, E. A., Nakashima, A., Takeuchi, H., Yamaguchi, S., et al. (2020). Visualization and molecular characterization of whole-brain vascular networks with capillary resolution. *Nat. Commun.* 11, 1104. doi: 10.1038/s41467-020-14786-z

Murphy, M. P., and LeVine, H. III. (2010). Alzheimer's disease and the amyloid-beta peptide. *J. Alzheimers Dis.* 19, 311–323. doi: 10.3233/JAD-2010-1221

Nobakht, M., Hoseini, S. M., Mortazavi, P., Sohrabi, I., Esmailzade, B., Rahbar Rooshandel, N., et al. (2011). Neuropathological changes in brain cortex and hippocampus in a rat model of Alzheimer's disease. *Iran. Biomed. J.* 15, 51–58.

Oakley, H., Cole, S. L., Logan, S., Maus, E., Shao, P., Craft, J., et al. (2006). Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* 26, 10129–10140. doi: 10.1523/JNEUROSCI.1202-06.2006

Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., et al. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Res. Rev.* 30, 25–48. doi: 10.1016/j.arr.2016.01.002

Ragan, T., Kadiri, L. R., Venkataraju, K. U., Bahlmann, K., Sutin, J., Taranda, J., et al. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat. Methods* 9, 255–258. doi: 10.1038/nmeth.1854

Robertson, R. T., Levine, S. T., Haynes, S. M., Gutierrez, P., Baratta, J. L., Tan, Z., et al. (2015). Use of labeled tomato lectin for imaging vasculature structures. *Histochem. Cell Biol.* 143, 225–234. doi: 10.1007/s00418-014-1301-3

Serrano-Pozo, A., Frosch, M. P., Masliah, E., and Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 1, a006189. doi: 10.1101/cshperspect.a006189

Valm, A. M., Oldenbourg, R., and Borisy, G. G. (2016). Multiplexed spectral imaging of 120 different fluorescent labels. *PLoS ONE* 11, e0158495. doi: 10.1371/journal.pone.0158495

Wessels, J. T., Hoffman, R. M., and Wouters, F. S. (2008). The use of transgenic fluorescent mouse strains, fluorescent protein coding vectors, and innovative imaging techniques in the life sciences. *Cytometry A* 73, 490–491. doi: 10.1002/cyto.a.20548

Whitesell, J. D., Buckley, A. R., Knox, J. E., Kuan, L., Graddis, N., Pelos, A., et al. (2019). Whole brain imaging reveals distinct spatial patterns of amyloid beta deposition in three mouse models of Alzheimer's disease. *J. Comp. Neurol.* 527, 2122–2145. doi: 10.1002/cne.24555

Wu, J., He, Y., Yang, Z., Guo, C., Luo, Q., Zhou, W., et al. (2014). 3D BrainCV: simultaneous visualization and analysis of cells and capillaries in a whole mouse brain with one-micron voxel resolution. *Neuroimage* 87, 199–208. doi: 10.1016/j.neuroimage.2013.10.036

Xiong, B., Li, A., Lou, Y., Chen, S., Long, B., Peng, J., et al. (2017). Precise cerebral vascular atlas in stereotaxic coordinates of whole mouse brain. *Front. Neuroanat.* 11, 128. doi: 10.3389/fnana.2017.00128

Youmans, K. L., Tai, L. M., Kanekiyo, T., Stine, W. B. Jr., Michon, S. C., et al. (2012). Intraneuronal Aβ detection in 5xFAD mice by a new Aβ-specific antibody. *Mol. Neurodegener.* 7, 8. doi: 10.1186/1750-1326-7-8

Yuan, J., Gong, H., Li, A., Li, X., Chen, S., Zeng, S., et al. (2015). Visible rodent brain-wide networks at single-neuron resolution. *Front. Neuroanat.* 9, 70. doi: 10.3389/fnana.2015.00070

Yuan, P., Condello, C., Keene, C. D., Wang, Y., Bird, T. D., Paul, S. M., et al. (2016). TREM2 haplodeficiency in mice and humans impairs the microglia barrier function leading to decreased amyloid compaction and severe axonal dystrophy. *Neuron* 90, 724–739. doi: 10.1016/j.neuron.2016.05.003

Zhang, J., Long, B., Li, A., Sun, J., Tian, J., Luo, T., et al. (2020). Whole-brain three-dimensional profiling reveals brain region specific axon vulnerability in 5xFAD mouse model. *Front. Neuroanat.* 14, 608177. doi: 10.3389/fnana.2020.608177

Zhang, X., Xianzhen, Y., Jingjing, Z., Anan, L., Hui, G., Qingming, L., et al. (2019). High-resolution mapping of brain vasculature and its impairment in the hippocampus of Alzheimer's disease mice. *National Science Review* 6, 1223–1238. doi: 10.1093/nsr/nwz124

Zhong, W., Gao, X., Wang, S., Han, K., Ema, M., Adams, S., et al. (2017). Prox1-GFP/Flt1-DsRed transgenic mice: an animal model for simultaneous live imaging of angiogenesis and lymphangiogenesis. *Angiogenesis* 20, 581–598. doi: 10.1007/s10456-017-9572-7

Zingg, B., Chou, X. L., Zhang, Z. G., Mesik, L., Liang, F., Tao, H. W., et al. (2017). AAV-Mediated anterograde transsynaptic tagging: mapping corticocollicular input-defined neural pathways for defense behaviors. *Neuron* 93, 33–47. doi: 10.1016/j.neuron.2016.11.045

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 Yin, Zhang, Zhang, Yang, Sun, Zhang, Gao and Jiang. 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 Neuroscience

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fnins.2022.850906

Robust Latent Multi-Source Adaptation for Encephalogram-Based Emotion Recognition

Jianwen Tao[1†], Yufang Dan[1†], Di Zhou[2†] and Songsong He[1†]*

[1] *Institute of Artificial Intelligence Application, Ningbo Polytechnic, Ningbo, China,* [2] *Industrial Technological Institute of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou, China*

In practical encephalogram (EEG)-based machine learning, different subjects can be represented by many different EEG patterns, which would, in some extent, degrade the performance of extant subject-independent classifiers obtained from cross-subjects datasets. To this end, in this paper, we present a robust Latent Multi-source Adaptation (LMA) framework for cross-subject/dataset emotion recognition with EEG signals by uncovering multiple domain-invariant latent subspaces. Specifically, by jointly aligning the statistical and semantic distribution discrepancies between each source and target pair, multiple domain-invariant classifiers can be trained collaboratively in a unified framework. This framework can fully utilize the correlated knowledge among multiple sources with a novel low-rank regularization term. Comprehensive experiments on DEAP and SEED datasets demonstrate the superior or comparable performance of LMA with the state of the art in the EEG-based emotion recognition.

Keywords: encephalogram, latent space, emotion recognition, co-adaptation, maximum mean discrepancy

OPEN ACCESS

Edited by:
Yuanpeng Zhang, Nantong University, China

Reviewed by:
Jian Zhang, Wuhan University of Technology, China
Fangfang Duan, Wuhan University of Technology, China

****Correspondence:***
Di Zhou
sion2005@sasu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *08 January 2022*
Accepted: *11 February 2022*
Published: *27 April 2022*

Citation:
Tao J, Dan Y, Zhou D and He S (2022) Robust Latent Multi-Source Adaptation for Encephalogram-Based Emotion Recognition. Front. Neurosci. 16:850906. doi: 10.3389/fnins.2022.850906

INTRODUCTION

Contemporarily, in the field of affective computing research, automated emotion recognition (AER) has attracted lots of attention from machine learning and computer vision (Kim et al., 2013). In traditional schema, one auto emotion recognition system driven by EEG signals usually includes two core components, i.e., feature extraction followed by emotion classification (Lan et al., 2018). Some representative EEG feature extraction methods (Jenke et al., 2014; Zhang et al., 2020b) can be viewed comprehensively in Jenke et al. (2014). This work mainly focuses on machine learning-based emotion classification methods.

In the past decade, a large scale of emotion recognition methods has been presented for effective emotion recognition using EEG features (Musha et al., 1997; Kim et al., 2013; Li et al., 2018b,c). Zheng (2017) proposed a novel emotion recognition method by exploiting the group sparse canonical correlation analysis, thus simultaneously implementing EEG channel selection and emotion recognition. Recently, Li et al. (2018c) also presented a sparse linear regression model with graph regularization for emotion recognition using EEG signals. In the past decade, due to their outperformed performance compared with traditional methods, deep emotion recognition methods using EEG signals have been widely explored in emotion feature extraction and recognition (Lotfi and Akbarzadeh-T, 2014), such as criminal psychological emotion recognition based on deep learning and EEG

signals (Liu and Liu, 2021), EEG-based Deep Belief Network model (Zheng and Lu, 2015), multi-channel EEG-based recognition model (Song et al., 2018), and EEG-based neural network model (Li et al., 2018b).

It is worthy to note that the aforementioned works for emotion recognition perform well only in such scenario that both training and test samples follow the same distribution (Zhang et al., 2020a), in which the recognition models obtained from the source dataset(s) therefore can be easily utilized in the target dataset effectively (Zhang et al., 2019a). Unfortunately, these traditional methods may fail in addressing cross-subject/dataset emotion recognition due to the mismatch of feature distribution with EEG signals. To address this issue, many domain adaptation (DA) emotion recognition models for AER problem have been promoted (Chu et al., 2017; Li et al., 2018a; Li et al., 2020a,b; Bao et al., 2021; Wang et al., 2021). In a DA emotion recognition system, one usually focuses on exploring an effective recognition model on one target domain with few or even none of the labeled data, by borrowing some positive knowledge from other source domain(s) with slightly different distribution with that of the target domain (Bruzzone and Marconcini, 2010; Tao et al., 2012; Long et al., 2014; Zhang et al., 2019b).

A typical challenge in one EEG-based emotion recognition system is the cross-subject/dataset learning problem (Li et al., 2018a). In such scenario, DA techniques can be exploited to address this challenging issue where both training and test data follow slightly different distribution (Tao et al., 2012; Long et al., 2014; Li et al., 2021). To deal with the challenging cross-subject EEG emotion recognition problem, Pandey and Seeja (2019) proposed a subject-independent approach for EEG emotion recognition. Li et al. (2018a) proposed another method for cross-subject EEG emotion recognition. In the past decade, deep neural networks (DNNs) (Ganin et al., 2016; Li et al., 2018a) have also driven rapid progress in DA (Duan et al., 2012a; Ding et al., 2018a). The DA issues can be solved by the domain adversarial neural network (DANN) (Ganin et al., 2016). It remains unclear, however, whether the performance of deep DA methods is really contributed by their deep feature representation, the fine-tuned classifiers, or is rather an outcome of the adaptation regularization terms (Ghifary et al., 2017).

Although existing DA methods have obvious effectiveness and efficiency in the special use of emotion recognition (Chu et al., 2017), there is few work to use the joint feature selection method and then carry out the multi-source adaptive domain recognition of cross datasets by exploiting the correlation knowledge among domains and features. Besides, during DA, most of the multi-source domain adaptation (MDA) methods (Yang et al., 2007; Duan et al., 2012b,c; Tommasi et al., 2014; Tao et al., 2015, 2017; Ding et al., 2018b) generally cope with the sources independently without considering the correlation information among the source domains (Zhang et al., 2019c), which may destroy the discriminant structure (either intrinsic or extrinsic) of multi-source domains (Rosenstein et al., 2005). Last but not the least, for an MDA system, it is crucial for source weight determination during learning based on the correlation and quality of source domains. To the best of our knowledge, these characters are not feasible enough in extant MDA methods.

In order to solve the above problems in existing MDA, we explore to exploit the relevant knowledge among sources in the uncovered subspaces to learn a multi-source adaptive emotion recognition model. In other words, we mainly adopt the strategy of digging the relationship between multi-source domains and one target domain (including feature and distribution) for promoting multi-source adaptive emotion recognition with EEG signals. We aim to progress beyond existing works that have partially addressed those issues by exploring to solve all the above-mentioned issues in a unified framework. Specifically, we propose in this work a robust Latent Multiple-source Adaption (LMA) method for EEG-based emotion recognition by mining multiple shared latent subspaces, each for one source–target domain pair. The method employs the robust regression scheme to process high-dimensional, sparse outliers and non-i.i.d. (independently identical distribution) EEG features by jointly utilizing the $l_{2,1}$-norm (Nie et al., 2010a) and trace norm. Under this framework, the row sparsity regularization is designed to obtain the solution of sparse feature selection (Zhang et al., 2020b). We match distributions between each domain pair (including both target and multi-source domains) by minimizing the nonparametric Maximum Mean Discrepancy (MMD) (Gretton et al., 2009; Pan et al., 2011) in each uncovered latent space shared by this source–target pair. The contributions of this paper are listed as follows:

(1) We propose a unified multi-source adaptive emotion recognition framework with EEG features by uncovering multiple latent subspaces.
(2) Our framework selects features in a collaborative way and considers the correlated knowledge among sources. In LMA, the importance of each feature does not need to be evaluated separately. In addition, in our unified framework, we can learn multiple loss functions with feature selection for all source adaptation subjects synchronously, so that our framework can use the correlated information of multiple sources as auxiliary information.
(3) In this framework, the original geometric structure is retained by using the graph Laplacian regularization, and the $l_{2,1}$-norm minimization sparse regression approach is used to suppress the influence of noise or outliers in the domains, which shows the robustness of the framework.
(4) Through a large number of experiments on two EEG datasets, we prove the effectiveness and convergence of this framework.

The remainder of the paper is organized as follows: In section Related Work, we discussed the related works with feature selection and multi-source DA learning. In section Proposed Framework, our framework LMA will be designed, and section Algorithm arranges the corresponding optimal algorithm of LMA. The experimental results and analysis on two real EEG datasets are presented in section Experimental Evaluation. Finally, we conclude in Section Conclusion.

TABLE 1 | Notations and descriptions.

Notations	Descriptions
N	Sample number of each source–target pair
d	Feature dimensionality number
χ	Sample/feature space
Γ	Label/prediction space
$a = [a_1, a_2, ..., a_d]^T \in \mathbb{R}^d$	Vector a
$A \in \mathbb{R}^{n \times d}$	Matrix A
$A_{i,j}$	The (i, j)th element of the matrix A
$A_{i,:}$ and $A_{:,j}$	The i/jth row/column vector of A
$(\cdot)^T$	Transpose operator
$tr(\cdot)$	Trace operator
$\langle A, B \rangle = tr(A^T B)$	The inner product of two matrices A and B
$\|\|a\|\|_p := \left(\sum_{i=1}^{d} \|a_i\|^p\right)^{1/p}$	The p-norm of a vector a
$\|\|A\|\|_{2,1} = \sum_{i=1}^{n} \|\|A_{i,:}\|\|_2 = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{d} A_{ij}^2}$	The $l_{2,1}$-norm of A
$\|\|A\|\|_* = tr((AA^T)^{\frac{1}{2}})$	The trace-norm of A
I_r	Identity matrix of size $r \times r$
1_d	d-dimensional vector with all ones
0_d	d-dimensional vector with all zeroes

RELATED WORK

In the past decades, affective computing community has paid increasing attention to the emotion recognition with brain–computer interfaces (BCI) (Mühl et al., 2014; Chu et al., 2017). A brain–computer interface system could capture certain emotion states and respectively make corresponding response to these states using spontaneous EEG signals even when explicit input from the subjects is unavailable (Zhang et al., 2019a), thus augmenting the user experience in the session of interactivity. Nowadays, a large number of methods (Zhang et al., 2016, 2017) have been proposed to recognize different emotion information from brain-wave signals. The latest works about affective BCI (aBCI) took account of machine learning algorithms on emotion recognition using a few discriminative features (Jenke et al., 2014; Mühl et al., 2014). In one representative BCI system, a certain feature extractor firstly extracts discriminative features from the raw EEG data, and then these features as well as labeled emotion states are sent into the classifier for real-time affection recognition. In the last decade, many aBCI-related works have presented sound and interesting emotion recognition performance (Mühl et al., 2014).

Although existing methods have obtained satisfied achievements on EEG-based emotion recognition, the expected performance could still be degraded by certain impacts in the case of cross-subject/dataset recognition due to the difference between subjects/datasets. Therefore, one needs to train a specific classifier for individual subject/dataset-of-interest. Even for the same subject, it is also indispensable to recalibrate the classifier frequently for maintaining a satisfied recognition accuracy since the EEG signals are unstable now and then. This would undoubtedly increase the costs of manual labor as well as time. Fortunately, the DA (a.k.a. domain transfer) technique can be leveraged to tackle these issues existing in EEG-based emotion recognition.

In the past decade, DA technique (Duan et al., 2012a,b,c; Tzeng et al., 2015; Tzeng et al., 2017; Ding et al., 2018a,b,c) has elicited an increasing attention in the community of machine learning. Up to now, domain-adaptation-based emotion recognition methods have nearly dominated the literature of aBCI Dolan, 2002; Mühl et al., 2014; Jayaram et al., 2016; Lan et al., 2018; Zhong et al., 2020; Zheng et al., 2015; Zheng and Lu, 2016; Chai et al., 2016; Chai et al., 2017; Shi et al., 2013; Koelstra et al., 2012; Zheng and Lu, 2015), which aim to address different issues in emotion classification by pursuing various DA skills using the EEG datasets such as SEED. In these preceding works, a commonly used strategy is to uncover a shared subspace from different domains by preserving certain discriminative properties, thus decreasing the differences among subjects or sessions extracted from the captured EEG signals (Tao and Dan, 2021). While extensive exploration on cross-subject/session has been conducted effectively in the prior works by leveraging various DA tricks, one obvious shortage in these works is that the evaluation dataset is just limited to one single database, e.g., SEED. In practical aBCI applications, the EEG datasets could also change since the EEG signals may be produced by different subjects, sessions, EEG devices, experimental schemes, and emotional stimuli. Henceforth, one of the yet unsolved issues in current research is the robustness and effectiveness of the proposed DA methods on cross-datasets/subjects.

PROPOSED FRAMEWORK

Notation

In the context, the symbol definitions are listed in **Table 1**. We respectively denote by $[A_1, A_2, ..., A_k]$ and $[A_1; A_2; ...; A_k]$ the concatenation of k matrices according to the row (horizontally) and the column (vertically). In this work, we focus on the multi-source adaptation framework, which can be driven by S source domains of c-class. We denote by $X^a = \{x_1^a, ..., x_{n_a}^a\} \in \mathbb{R}^{d \times n_a}$ $(a = 1, 2, ..., S)$ the ath source dataset with n_a samples[1]. Its corresponding class label matrix can be denoted as $Y^a = [y_1^a, ..., y_{n_a}^a]^T \in \mathbb{R}^{n_a \times c} \in \Gamma = \{0, 1\}^{c \times 1}$ with $y_{il} = 1$ if the ith sample is labeled as the lth class and -1 others. Correspondingly, we denote by $X^t = \{x_1^t, x_2^t, ..., x_m^t\} \in \mathbb{R}^{d \times n_t}$ the target dataset of interest. Since the true classes of the samples in X^t are inaccessible in the training stage, the target labels (or pseudo labels) $Y^t = [y_1^t, ..., y_{n_t}^t]^T \in \mathbb{R}^{n_t \times c} \in \Gamma$ can be predicted by certain pre-trained classifiers trained on the source datasets with labeled data. Therefore, detecting the ground-truth label of each target sample is our ultimate goal.

We further denote X^a/X^t with the label $\bar{l}$ as $X^{a(\bar{l})}/X^{t(\bar{l})}$ $(\bar{l} = 1, ..., c)$, and the ath source–target domain pair as $X_a = [X^a, X^t] \in \mathbb{R}^{d \times N}$ $(N = n_t + n_a)$ with label matrix $Y_a = [Y^a, Y^t]$ by packing the ath source and the target data.

[1]While we do not need to limit the number of instances in each source domain, which is identical with that assumed when shaped into the training matrix, for the sake of simplicity, we can extract the same number of training instances from each source domain.

Problem Statement

A commonly used strategy in the representative MDA is to acquire knowledge from multiple sources by leveraging certain common knowledge shared by them to promote the target learning of interest. We propose in this work a robust Latent Multiple-source Adaption (LMA) emotion recognition method based on EEG features. The method employs the robust regression scheme to process high-dimensional, sparse, outliers, and non-i.i.d. EEG features by jointly utilizing the $l_{2,1}$-norm and trace norm (Yang et al., 2013). The designed method has three characteristics, which are integrated into a unified optimization formulation to find an effective emotion recognition model by aligning the feature distribution between each source–target domain pair. Specifically, it includes four technical aspects: (1) *via* employing the $l_{2,1}$-norm minimization, a robust loss term is introduced into each source model learning by taking account of the influence of noise or outliers in EEG signal (Li et al., 2015), and a sparse regularization term is designed to eliminate over-fitting and a sparse features subset is selected; (2) based on the designed regression model and the semantic distribution matching between each pair of domains in each uncovered latent spaces (Tao et al., 2019), it not only provides robustness on loss function, but also retains the domain distribution (including local and global) structures (Nie et al., 2010b), and meanwhile maintains a high dependence on the (pseudo) label knowledge of the source domains and the target domain (Nie et al., 2010b; Ding et al., 2018c; Zhang et al., 2020a), so as to obtain preferable generalization performance; and (3) by exploiting the trace norm of matrix, we can make full use of the correlative information among multiple sources and transfer more discriminative knowledge to the target domain.

Specifically, we present the flow diagram of LWA in **Figure 1** to illustrate our innovation: firstly, we can project each source EEG data into one domain-invariant subspace by minimizing the domain-wise distribution discrepancy; thus, S classifiers are being jointly learned by employing trace norm as well as $l_{2,1}$-norm; we then obtain S target label matrices predicted from these source classifiers on the target domain; furthermore, in the original space, we also learn a target model using the squared regression scheme with the constraint of prediction consistency on the target data between those source models and the target model; and by uncovering multiple domain-invariant latent spaces, we finally formulate a joint learning framework of multi-source adaptation for EEG-based emotion recognition. To implement these properties, in the following part, we will detail the objective formulation of the proposed method.

General Formulation

In this section, we propose the general formulation of LMA framework underpinned by the robust regression principle and the regularization theory. We investigate the learning problem under multiclass setting, with the decision classifiers $\left\{f_\theta^a(x)\right\}_{a=1}^S$, where θ is the parameter of the hypothesis space of those decision functions. We propose a unified MDA framework by uncovering S discriminative latent subspaces Θ_a $(a = 1, 2, ..., S)$ (Tao and Xu, 2019) and to learn decision classifiers $f_\theta^a(x)$s of all sources simultaneously. In particular, the proposed method minimizes the distribution difference of each domain pair after the projection P_a into the subspace Θ_a, as well as the structural risk functional of the labeled data from the source domain X^a. We also let Θ_a be orthogonal on rows so that $\Theta_a\Theta_a^T = I_{r\times r}$, where $r(\ll d)$ is the dimensionality of the shared latent space. We then endeavor to find S cross-domain models parameterized by $\{\theta_a\}_{a=1}^S$ *via* jointly utilizing correlated knowledge among sources in some latent spaces. We therefore propose the following general formulation of LMA:

$$\begin{aligned}&\Re\left(f_{\theta_a}^a, f_{\theta_t}^t, \Theta_a\right)\\&=\sum_{a=1}^{S}\left(R(f_{\theta_a}^a, Y_a)+dist_{\Theta_a}(X^a, X^t)+\mho(f_{\theta_a}^a)\right)+Con(f_{\theta_a}^a, f_{\theta_t}^t, X^t)\\&\qquad+R(f_{\theta_t}^t, Y^t)+\Omega(f_\theta^a),\end{aligned}\tag{1}$$

where $\theta_{a/t}$ is the parameter set of the ath source/target model, $Con(f_{\theta_a}^a, f_{\theta_t}^t, X^t)$ enforces the discriminative consistency between the ath source and the target model on the target dataset, $R(\cdot,\cdot)$ is the robust regression model, the regularization term $\mho(f_{\theta_a}^a)$ controls the complexity of $f_{\theta_a}^a$, $dist_{\Theta_a}(X^a, X^t)$ is for aligning the distributions of each domain pair in the latent space Θ_a, the regularizer $\Omega(f_\theta^a)$ controls the low rankness of all source models for mining the correlated knowledge. Hence, by solving the objective in (1), the subspace Θ_a and the decision functions $f_{\theta_{a/t}}^{a/t}$s can be learned simultaneously. In the sequence, we will focus on designing these components in the general formulation one by one to construct a unified framework.

Design of Regression Model

In LMA, we learn a composite source classifier function $f^a = P_a^\circ\Theta_a$ trained on the EEG features, where Q_a represents the source classifier model, and $\circ$ is the function combination operator. We therefore explore to find the best approximation W_t for f^t by leverage Q_a in Θ_a with the assumption that there exist some commonalities (e.g., discriminative structures) between different domains (Tao et al., 2019). Moreover, it should also maintain the discriminative structure in the original space. To capture the source correlation information, we respectively design the following classification functions for the ath source domain in the latent space:

$$f_\theta^a(X^a) = V_a^T\phi(X^a) + Q_a^T\psi_\theta(X^a),\quad a = 1, 2, ..., S,\tag{2}$$

where $\phi : \chi \to H^2$ is a known feature map projecting the ath source data from the input space χ into certain reproducing kernel Hilbert space (RKHS) (Nie et al., 2010b) H. The other component ψ_θ is a parameterized low-dimensional space that aims at encoding the shared structure between each source domain and the target domain. The weight vector Q_a is defined in the projected subspace under the projection ψ_θ for the ath source, and V_a is the weight matrix defined in the original feature space. With the parametric form in (2), the learned subspace

[2]It is important to note that the feature mapping function ϕ with respect to each source domain can be completely different from each other.

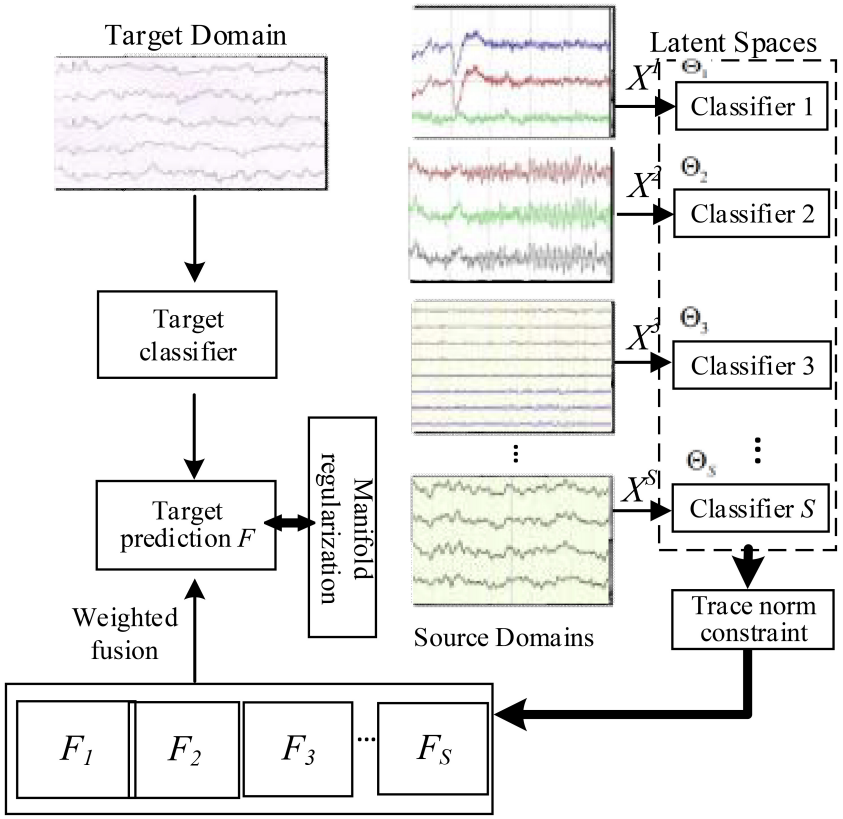


FIGURE 1 | Flowchart of LMA on EEG-based emotion recognition.

ψ_θ can capture the intrinsic structure of source correlation in the MDA problem, which are shared by each source and the target domain. Correspondingly, we also can design the target recognition model: $f^t(X^t) = W_t^T \phi(X^t)$ with the weight matrix W_t. We present the empirical kernel map as discussed in Gretton et al. (2009):

$$\psi_e : \quad \chi \to R^N, \ \textit{for linear kernel mapping}$$
$$x \to K_\psi(\cdot, x)|_{x_1,x_2,\ldots,x_N} = \left(K_\psi(x_1, x), \ldots, K_\psi(x_N, x)\right), \ . \ \textit{for nonlinear kernel mapping}$$

In the following, we will discuss both the linear classification function and the nonlinear (kernel) classification function and integrate them into a unified form.

- *Linear classifier*. We can consider a simple linear form of feature map, where $\theta = \Theta_a$ is an $r \times d$ dimensional matrix and $\psi_\theta(x) = \Theta_a \psi(x)$, with a known d-dimensional vector function $\psi(x)$. Furthermore, following (11), we can take a simple model $\phi(X^a) = \psi(X^a) = X^a$ into account. We can thereby write the linear classifier as

$$f_\theta^a(X^a) = V_a^T X^a + Q_a^T \Theta_a X^a, \quad a = 1, 2, \ldots, S. \tag{3}$$

- *Nonlinear classifier*. If we take kernel learning into account and assume that the feature map $\phi(x)$ and $\psi(x)$ belong to certain reproducing kernel Hilbert space (RKHS), Eq. (3) therefore can be kernelized. For $\psi(x)$, we firstly denote the kernel matrix as $K_\psi = \langle \psi(x_i), \psi(x_j) \rangle$. By using empirical kernel, we have kernel matrix $K^a = \phi(X^a)$ with $(K^a)_{i,j} = \left\langle \phi(x_i^a), \phi(x_j^a) \right\rangle$, where $x_i^a, x_j^a \in X^a$. Finally, we can let $\psi_\theta = \Theta_a \psi_e$, where $\Theta_a \in R^{r \times N}$ is used to transform the empirical kernel vector to an r-dimensional space. Let $\{\Psi_a^i\}_{i=1}^r$ denote the weight parameters in the embedded kernel subspace for the ath source. Hence, the kernelized decision functions become

$$f_\theta^a(x) = \omega_a^T K^a(\cdot, x) + \Psi_a^T \Theta_a K^a(\cdot, x). \tag{4}$$

where ω_a is the weight coefficients in the original kernel space for the ath source.

In order to model the linear case in Eq. (3) and kernel case in Eq. (4) into a unified framework, we introduce

$$T_a = \begin{cases} V_a + Q_a^T \Theta_a, & \textit{linear} \\ \omega_a + \Psi_a^T \Theta_a & \textit{kernel} \end{cases}. \tag{5}$$

Moreover, in the following, we use two symbols, namely, W_a and P_a, where W_a denotes V_a in the linear case and ω_a in the kernel case, and P_a denotes Q_a in the linear case and Ψ_a in the kernel case. Then, Eq. (5) becomes $T_a = W_a + P_a^T \Theta_a$ for both linear and

kernel cases, and we can represent the data in linear space and nonlinear space as follows:

$$\overline{X}^{a/t} = \begin{cases} X^{a/t}, \ linear \\ K^{a/t}(\cdot, x) \ , \ kernel \end{cases}. \tag{6}$$

In the sequence, we also refer to $X^{a/t}$ as $\overline{X}^{a/t}$ if without special denotation for simplicity of expression. As a result, we can formulate the predictors, linear form as in Eq. (4) and nonlinear form as in Eq. (6), in a unified form as depicted in

$$f_\theta^a(X^a) = T_a^T X^a, \quad a = 1, 2, ..., S \tag{7}$$

We introduce the sparse regression scheme (Shi et al., 2015) by exploiting $l_{2,1}$-norm minimization to enhance the robustness against the misclassification. We particularly construct a scaled pseudo label matrix for the target data, i.e., $F = [f_1, f_2, ..., f_{n_t}] = (Y^t(Y^t)^T)^{-1/2}Y^t \in \mathbb{R}^{n_t \times c}$, where the scaled pseudo label $f_i = y_i^t$ if x_i^t is labeled, $f_i = 0$ otherwise. Therefore, $FF^T = I_{n_t}$ can be easily derived with additional constraint $F \geq 0$. We then respectively find the source classifiers trained on X^a $(a = 1, 2, ..., S)$ and the target classifier trained on X^t by minimizing the following loss functions.

$$R(f_{\theta_a}^a, Y_a) = \left|\left| f_{\theta_a}^a(X^a) - Y^a \right|\right|_{2,1}. \tag{8}$$

$$R(f^t, F) = \left|\left| f^t(X^t) - F \right|\right|_F^2 + \beta \left(||W_t||_{2,1} + tr(FLF^T) \right). \tag{9}$$

where L denotes the graph Laplacian matrix induced from the target samples.

Moreover, it is intuitively reasonable that the outputs of f^a s on the target domain are expected to be consistent with those of f^t, which would gradually make P_a and W_t more accurate after lots of iterations. This prediction consistency can be minimized *via* the following residual:

$$\begin{aligned} Con(f^a, f^t, X^t) = & \left|\left| X_t^T(W_a + \Theta_a P_a) - F^a \right|\right|_F^2 \\ & + \left|\left| X_t^T W_t - F \right|\right|_F^2 + \left|\left| F^a - F \right|\right|_F^2. \end{aligned} \tag{10}$$

In such a way, P_a and W_t would jointly enhance the target discriminations for the final emotion recognition.

Additionally, based on the parametric form of the decision function f_θ^a as in Eq. (7), we introduce the following regularizer:

$$\mho(f_\theta^a) = ||W_a||^2 + ||T_a||_{2,1} = \left|\left| T_a - \Theta_a^T P_a \right|\right|_F^2 + ||T_a||_{2,1}, \tag{11}$$

which controls the complexity of each source classifier independently in the original and latent subspaces, respectively.

Uncovering Latent Spaces

In this subsection, we will present an effective strategy to capture multiple domain-invariant subspaces to mitigate the domain discrepancy as well as excavate some domain-invariant discriminative information. To this end, we give two main constraints or conditions on uncovering these latent spaces: (1) preserving the within-domain local structures and (2) aligning the inter-domain marginal distribution divergence as well as conditional distribution discrepancy. Following existing feature extraction methods (Tao et al., 2016), we further constrain Θ_a to be orthogonal on rows, i.e., $\Theta_a \Theta_a^T = I_r$, where r (typically far less than d) is the feature dimensionality in the latent space.

To fulfill the first condition, we construct a locality preserving regularizer to measure the smoothness along the intrinsic discriminative structure of the domain features (Nie et al., 2010b; Shi et al., 2015; Ding et al., 2018c). Specifically, one can construct an undirected graph with a weighted adjacency matrix $\prod_a = [(\prod_a)_{i,j}]_{i,j=1,2,...,N}$, which is defined as (Yan et al., 2006):

$$(\prod_a)_{i,j} = \begin{cases} \exp(-\gamma||x_i - x_j||^2), \ \text{if } x_i \in \delta_k(x_j) \text{ or } x_j \in \delta_k(x_i) \\ \quad \text{and both have the same labels} \\ \exp(-\frac{\gamma}{||x_i - x_j||^2}), \ \text{if } x_i \in \delta_k(x_j) \text{ or } x_j \in \delta_k(x_i) \\ \quad \text{and both have different labels,} \\ 0, \ \text{otherwise} \end{cases}, \tag{12}$$

where $x_i, x_j \in X_a = [X^a, X^t]$, $\delta_k(x)$ denotes the k nearest neighbor set of x, and the hyper-parameter γ can be empirically computed as $\overline{\theta}_a\sqrt{c}$ due to the impact of multi-class distribution, where $\overline{\theta}_a$ is the squared root of the mean norm of X_a. Deriving a diagonal matrix Δ_a from $\prod_a$ with $(\Delta_a)_{i,i} = \sum_j (\prod_a)_{i,j}$, we then compute the graph Laplacian matrix as $L_a = \Delta_a - \prod_a$. Thus, preserving the local geometrical structures of EEG features can be implemented by the following commonly used formulation in the manifold learning (Chen et al., 2013).

$$tr(\Theta_a^T X_a L_a X_a^T \Theta_a). \tag{13}$$

Benefiting from its simplicity and effectiveness, Maximum Mean Discrepancy (Gretton et al., 2009; Pan et al., 2011) has been commonly used to measure the distribution distance between two different domains. Consequently, to meet the second condition, we aim to minimize the MMD in certain optimized RKHS (Gretton et al., 2009). Specifically, the MMD between each domain pair is defined as follows:

$$\begin{aligned} & MMD(X^a, X^t) \\ & = \sup_{||\phi|| \leq 1} \left(E_{X^a \sim P}[\phi(X^a)] - E_{X^t \sim Q}\left[\phi(X^t)\right] \right), \\ & = \left|\left| E_{X^a \sim P}[\phi(X^a)] - E_{X^t \sim Q}\left[\phi(X^t)\right] \right|\right|_H \end{aligned} \tag{14}$$

The empirical counterpart of the MMD in Eq. (14) can be defined as:

$$MMD(X^a, X^t) = \left|\left| \frac{1}{n_a} \sum_{x_i \in X^a} \phi(x_i) - \frac{1}{n_t} \sum_{x_j \in X^t} \phi(x_j) \right|\right|_H, \tag{15}$$

which can recover an asymptotically unbiased estimation of the squared MMD in Eq. (14). Denote the gram matrix $\widetilde{K}_a(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ on dataset X_a as

$$\widetilde{K}_a = \begin{bmatrix} \widetilde{K}^a & \widetilde{K}^{at} \\ \widetilde{K}^{ta} & \widetilde{K}^t \end{bmatrix} \in R^{N \times N}, \tag{16}$$

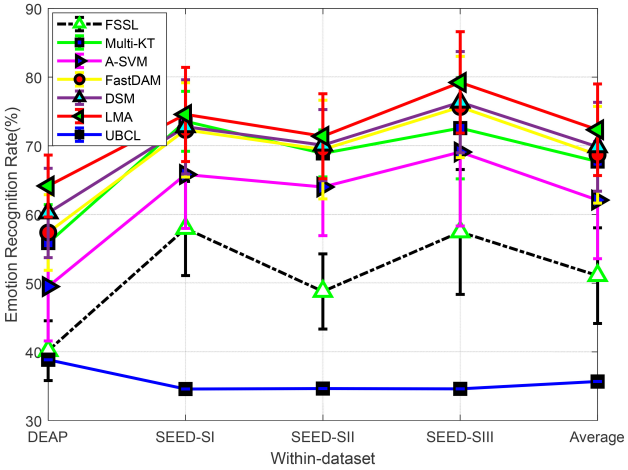


FIGURE 2 | Domain adaptation emotion recognition on within-dataset (SI: Session I, SII: Session II, SIII: Session III).

where $\widetilde{K}^a$, $\widetilde{K}^t$, and $\widetilde{K}^{at}$ (or $\widetilde{K}^{ta}$) are the Gram matrices respectively defined on the source domain, target domain, and cross domain data. Thus, the squared MMD in Eq. (15) can be formulated as

$$MMD(X^a, X^t) = tr\left(\widetilde{K}_a D^a\right), \tag{17}$$

where

$$D_{ij}{}^a = \begin{cases} \frac{1}{n_a^2} & when\ x_i, x_j \in X^a \\ \frac{1}{n_t^2} & when\ x_i, x_j \in X^t \\ \frac{-1}{n_a n_t} & otherwise \end{cases}. \tag{18}$$

In the sequence, we will take into account the feature map ϕ in linear as well as kernel forms:

λ Linear kernel: $\widetilde{K}_a = X_a^T \Theta_a^T \Theta_a X_a$ if $\phi(x) = \Theta_a x$, where $X_a = [x_1, x_2, ..., x_N]$.

λ Nonlinear kernel: $\widetilde{K}_a = K_a^T \Theta_a^T \Theta_a K_a$ if $\phi(x) = \Theta_a \psi_e(x) = \Theta_a K_a(\cdot, x)$, where $\psi_e(x)$ is the empirical kernel map defined in Eq. (4).

Recalling the definition of $\overline{X}$ in Eq. (6), the domain discrepancy criterion defined in (17) can be reformulated:

$$MMD^2(X^a, X^t) = tr(\widetilde{K}_a D^a) = tr(\Theta_a^T \overline{K}_a \Theta_a), \tag{19}$$

where

$$\overline{K}_a = \begin{cases} X_a D^a X_a^T, & linear\ kernel \\ K_a^T D^a K_a, & nonlinear\ kernel \end{cases}. \tag{20}$$

Note that Eq. (19) could not preserve the local structures of the EEG data from the same class in the latent spaces due to the shortage of semantic alignment. This would significantly deteriorate the learning performance in some cases. To this end, we further address this issue by improving Eq. (19) with the following class distribution matching term:

$$\sum_{l=1}^{c} \left\| \frac{1}{n_a^l} \sum_{i=1}^{n_a^l} \Theta_a^T \phi(x_i^{a(c)}) - \frac{1}{n_t^l} \sum_{j=1}^{n_t^l} \Theta^T \phi(x_j^{t(c)}) \right\|_F^2$$

$$= \sum_{l=1}^{c} tr(\Theta_a^T \overline{K}_a^{(l)} \Theta_a), \tag{21}$$

where

$$\overline{K}_a^l = \begin{cases} X_a^{(l)} D^{a(l)} (X_a^{(l)})^T, & linear \\ (K_a^{(l)})^T D^{a(l)} K_a^{(l)}, & kernel \end{cases},$$

$X_a^{(l)} = [X^{a(l)}, X^{t(l)}]$, with $X^{a(l)}$ and $X^{t(l)}$ being the datasets of the lth class, respectively, from source and target domains, n_a^l (respectively n_t^l) is the data size of the lth class from the ath source (respectively target) domain, and the elements of the matrix $D^{a(l)} = [Di, j^{a(l)}]$ are defined as

$$D_{i,j}^{a(l)} = \begin{cases} \frac{1}{(n_a^l)^2}, & when\ x_i, x_j \in X^{a(l)} \\ \frac{1}{(n_t^l)^2}, & when\ x_i, x_j \in X^{t(l)} \\ \frac{-1}{n_a^l n_t^l}, & otherwise \end{cases}. \tag{22}$$

Equation (21) explicitly forces EEG data from different domains but the same class to be mapped adjacently in the latent spaces. By unifying Eq. (19) and Eq. (21), we can obtain:

$$\sum_{l=0}^{c} tr(\Theta_a^T \overline{K}_a^{(l)} \Theta_a) = tr\left(\Theta_a^T \left(\sum_{l=0}^{c} \overline{K}_a^{(l)}\right) \Theta_a\right), \tag{23}$$

where $\overline{K}a^{(0)} = \overline{K}a$ and $D^{a(0)} = D^a$. We further denote $\Lambda^a = \sum_{l=0}^{c} \overline{K}_a^{(l)}$. By combining Eq. (13) and Eq. (23), we can attempt to uncover a latent space by minimizing the following formulation:

$$dist_{\Theta_a}(X^a, X^t) = tr\left(\Theta_a^T (X_a L_a X_a^T + \Lambda^a) \Theta_a\right)\ \ s.t.\ \Theta_a^T \Theta_a = I_r. \tag{24}$$

Sharing Source Discriminative Structure

While each source model is learned in different latent space from each other, one still can presume that these source models might be correlated due to the correlation of source EEG signals in the

model level (Tao et al., 2016). These correlated discriminative structures can be encoded by a low rank matrix of all source models, thus transferring the source knowledge from each other. For its simplicity of computation, the following trace norm of the matrix $P = [P_1, P_2, ..., P_S]$, which is a surrogate of the rank minimization, can be adopted for correlating the source models.

$$\Omega(f_\theta^a) = ||P||_* = tr\left(\left(PP^T\right)^{\frac{1}{2}}\right), \tag{25}$$

Overall Formulation

Combining the above formulations respectively defined in Eqs (8) to (11) and Eqs (24) and (25) together, we have the following objective function:

$$\begin{aligned}\Re = \underset{\substack{P_a, W^t, T_a, F_a,\\ F, W^t, \vartheta, \eta, \Theta}}{\arg\min} & \sum_{a=1}^{S}\left[\vartheta_a^{q_1}\left|Y_a - X_a^T T_a\right|_{2,1}\right. \\ & \left.+\alpha\left(|T_a|_{2,1} + |T_a - \Theta_a P_a|_F^2\right) + \eta_a^{q_2}\left|F_a - X_t^T T_a\right|_{2,1}^2\right], \\ & + \sum_{a=1}^{S}\eta_a^{q_2} tr(\Theta_a^T C_a \Theta_a) + \sum_{a=1}^{S}\eta_a^{q_2}|F_a - F|_F^2 \\ & + \left|X_t^T W_t - F\right|_F^2 + \beta\left(|W_t|_{2,1} + tr(FLF^T)\right) + \frac{\lambda}{2}|P|_* \\ s.t.\ & \sum_{a=1}^{S}\vartheta_a = \sum_{a=1}^{S}\eta_a = 1, \Theta_a^T\Theta_a = I_r\end{aligned} \tag{26}$$

where $C_a = X_a L_a X_a^T + \Lambda^a$, $\vartheta = [\vartheta_1, ..., \vartheta_a]^T$ is the weight vector to jointly combine all source regression loss, α, β, and λ are three regularization parameters, $q_1, q_2 > 1$ are two tunable parameters for avoiding trivial solution, and the tunable vector $\eta = [\eta_1, \eta_2, ..., \eta_S]$ denotes the adaptation degrees of different sources. The $l_{2,1}$-norm regularization added on projection matrix P_a forces most of the rows in P_a ($a = 1, ..., S$) to shrink to zero, thus performing feature selection on original data.

Emotion Recognition

After the best model parameters have been pursued, the source and target classifiers can be applied to recognize the emotion level of each probe EEG data. Specifically, we linearly fuse two recognition results on the probe data, i.e., $f^s(X^t) = \sum_{a=1}^{S}\vartheta_a (X^t)^T(\Theta_a P_a + W_a)$ obtained from source models, and $f^t(X^t) = (X^t)^T W^t$ predicted from the target model, as the final prediction value. That is, the following combination function can be exploited for recognizing the emotion level of the given test data x_i^t:

$$j = \arg\max_j \left(y_i^t = \delta f^s(x_i^t) + (1-\delta) f^t(x_i^t)\right)_j,$$

where δ is a trade-off parameter, tuned from $[0, 1]$. In the experimental setting, we empirically set $\delta = 0.5$ for initialization, followed by the evaluation of its impact on performance with different values of it.

ALGORITHM

In this section, an alternately iterative procedure is adopted to optimize the objective function in Eq. (26), which is followed by an overall algorithm.

Optimization

In terms of the definition in Nie et al. (2010a), we can derive $\left|\widetilde{T}\right|_{2,1} = 2tr\left(\widetilde{T}^T Q \widetilde{T}\right)$, where Q is a diagonal matrix with the ith diagonal element $Q_{ii} = \frac{1}{2||\widetilde{T}_{i,:}||_2}$. Hence, we can further transform Eq. (26) into Eq. (27):

$$\begin{aligned}\Re = \underset{\substack{P_a, W_t, T_a, F_a,\\ F, \vartheta, \eta, \Theta}}{\arg\min} & \sum_{a=1}^{S}\begin{pmatrix}\vartheta_a^{q_1} tr(\widetilde{T}_a^T Z_a \widetilde{T}_a) + \alpha tr\left(T_a^T G_a T_a\right. \\ \left.+|T_a - \Theta_a P_a|_F^2\right) \\ +\eta_a^{q_2}\left(tr(\Theta_a^T C_a \Theta_a) + |F_a - F|_F^2\right. \\ \left.+tr(Q_a^T \widetilde{Z}_a Q_a)\right)\end{pmatrix}, \\ & + \frac{\lambda}{2} tr\left(P^T\left(PP^T\right)^{-\frac{1}{2}}P\right) + \left|X_t^T W_t - F\right|_F^2 + \beta\left(|W_t|_{2,1} + tr(FLF^T)\right) \\ s.t.\ & \sum_{a=1}^{S}\vartheta_a = \sum_{a=1}^{S}\eta_a = 1, \Theta_a^T\Theta_a = I_r\end{aligned} \tag{27}$$

where $\widetilde{T}_a = (Y_a - X_a^T T_a)$ and $Q_a = F_a - X_t^T T_a$.

By solving the derivative of Eq. (27) w.r.t. W_t and letting it equal to zero:

$$\frac{\partial \Re}{\partial W_t} = 0 \Rightarrow W_t = \Delta^{-1} X_t F, \tag{28}$$

where $\Delta = X_t X_t^T + \beta\widetilde{V}$, and $\widetilde{V}$ is a diagonal matrix with the kth diagonal element being $(\widetilde{V})_{kk} = \frac{1}{2||(W_t)_{k,:}||_2}$. Substituting W_t in Eq. (27) with Eq. (28), we have:

$$\begin{aligned}\Re = \underset{\substack{P_a, T_a, F_a,\\ F, \vartheta, \eta, \Theta}}{\arg\min} & \sum_{a=1}^{S}\begin{pmatrix}\vartheta_a^{q_1} tr(\widetilde{T}_a^T Z_a \widetilde{T}_a) + \alpha\left(tr\left(T_a^T G_a T_a\right)\right. \\ \left.+|T_a - \Theta_a P_a|_F^2\right) + |F_a - F|_F^2 \\ +\eta_a^{q_2}\left(tr(\Theta_a^T C_a \Theta_a) + tr((F_a - X_t^T T_a)\right. \\ \left.^T \widetilde{Z}_a (F_a - X_t^T T_a))\right)\end{pmatrix}, \\ & + \frac{\lambda}{2} tr\left(P^T\left(PP^T\right)^{-\frac{1}{2}}P\right) + \beta tr\left(F\widetilde{\Delta}F^T\right) \\ s.t.\ & \textstyle\sum_{a=1}^{S}\vartheta_a = \sum_{a=1}^{S}\eta_a = 1, \Theta_a^T\Theta_a = I_r\end{aligned} \tag{29}$$

where $\widetilde{\Delta} = \beta L + \left(X_t^T \Delta^{-1} X_t - I\right) + \beta X_t^T \Delta^{-1} \widetilde{V} \Delta^{-1} X_t$. By solving the derivative of Eq. (29) w.r.t. F_a and letting it equal to zero, we have:

$$\begin{aligned}\frac{\partial \Re}{\partial F_a} &= 2\widetilde{Z}_a F_a - 2\widetilde{Z}_a X_t^T T_a + 2F_a - 2F = 0 \\ &\Rightarrow F_a = S_a^{-1}\left(F + \widetilde{Z}_a X_t^T T_a\right)\end{aligned}, \tag{30}$$

where $S_a = \widetilde{Z}_a + I$. Plugging F_a in Eq. (30) into Eq. (29), we can get:

$$\begin{aligned}&\eta_a^r tr\left(\left(S_a^{-1}F + \left(S_a^{-1}\widetilde{Z}_a X_t^T - X_t^T\right)T_a\right)^T \widetilde{Z}_a \left(S_a^{-1}F + \left(S_a^{-1}\widetilde{Z}_a X_t^T - X_t^T\right)T_a\right)\right) \\ &+\eta_a^r \left|\left|(S_a^{-1} - I)F + S_a^{-1}\widetilde{Z}_a X_t^T T_a\right|\right|_F^2 + \beta tr\left(F\widetilde{\Delta}F^T\right)\end{aligned}. \tag{31}$$

By solving the derivative of Eq. (31) w.r.t. F and letting it equal to 0, we obtain:

$$F = G_a^{-1} D_a T_a, \tag{32}$$

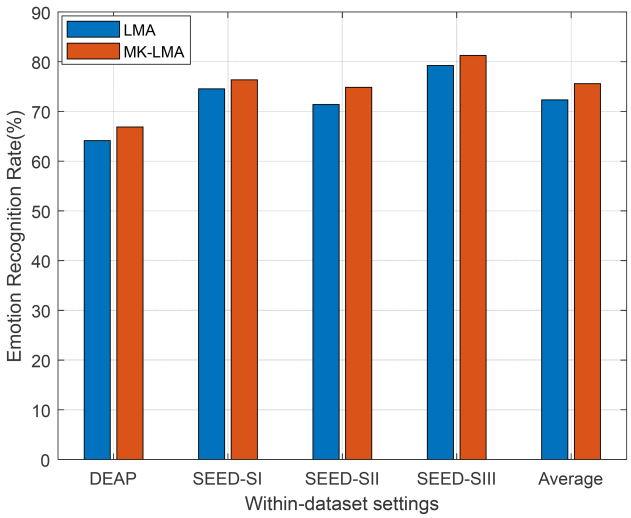


FIGURE 3 | Domain adaptation emotion recognition on within-dataset with multi-kernel learning (SI: Session I, SII: Session II, SIII: Session III).

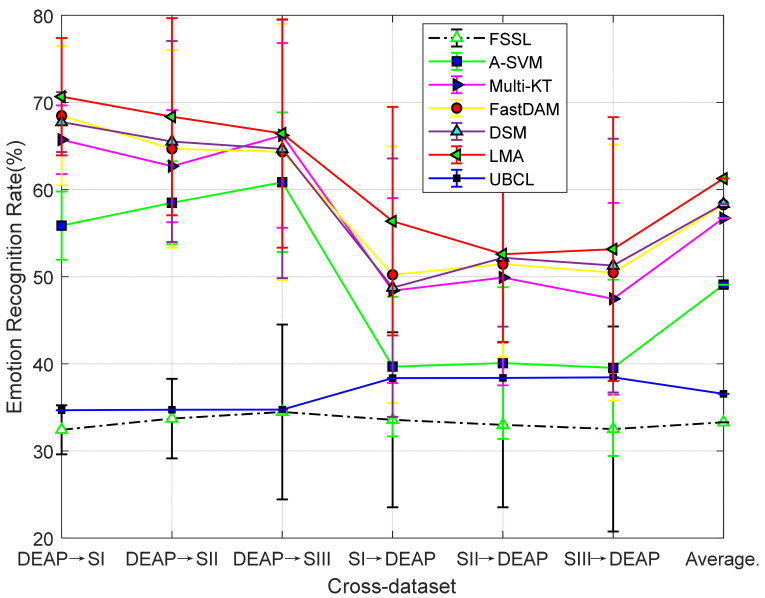


FIGURE 4 | Domain adaptation emotion recognition on cross-dataset (SI: Session I, SII: Session II, SIII: Session III).

where

$$\begin{cases} G_a = \eta_a^{q_2} S_a^{-1} \widetilde{Z}_v S_a^{-1} + \eta_a^{q_2} (S_a^{-1} - I)^T (S_a^{-1} - I) + \widetilde{\Delta} \\ D_a = \eta_a^{q_2} \left(S_a^{-1} - I\right) S_a^{-1} \widetilde{Z}_a X_t^T + \eta_a^{q_2} S_a^{-1} \widetilde{Z}_a \left(S_a^{-1} \widetilde{Z}_a - I\right) X_t^T \end{cases}, \tag{33}$$

By replacing F and F_a in Eq. (29) with those in Eqs (30) and (32), respectively, and solving the derivative of Eq. (29) with reference to T_a and equaling to zero, we then get:

$$T_a = H_a^{-1} \left(\vartheta_a^{q_1} X_a Z_a Y_a + \alpha \Theta_a P_a\right), \tag{34}$$

where:

$$\begin{aligned} H_a \; = \; & \vartheta_a^{q_1} X_a Z_a X_a^T + \alpha \left(G_a + I\right) + \eta_a^{q_2} \left((S_a^{-1} - I) G_a^{-1} D_a \right. \\ & + S_a^{-1} \widetilde{Z}_a X_t^T \left.\right)^T \left((S_a^{-1} - I) G_a^{-1} D_a + S_a^{-1} \widetilde{Z}_a X_t^T\right) \\ & + \beta D_a^T G_a^{-1} \widetilde{\Delta} G_a^{-1} D_a + \eta_a^{q_2} \left(S_a^{-1} G_v^{-1} D_v + S_a^{-1} \widetilde{Z}_a X_t^T - X_t^T\right)^T \cdot \\ & \widetilde{Z}_a \left(S_a^{-1} G_v^{-1} D_v T_v + S_a^{-1} \widetilde{Z}_a X_t^T - X_t^T\right) \end{aligned} \tag{35}$$

Let $U_a = \frac{1}{2} \left(PP^T\right)^{-\frac{1}{2}}$ and by replacing T_a in Eq. (29) with Eq. (34), and solving the derivative of Eq. (29) in reference to P_a and

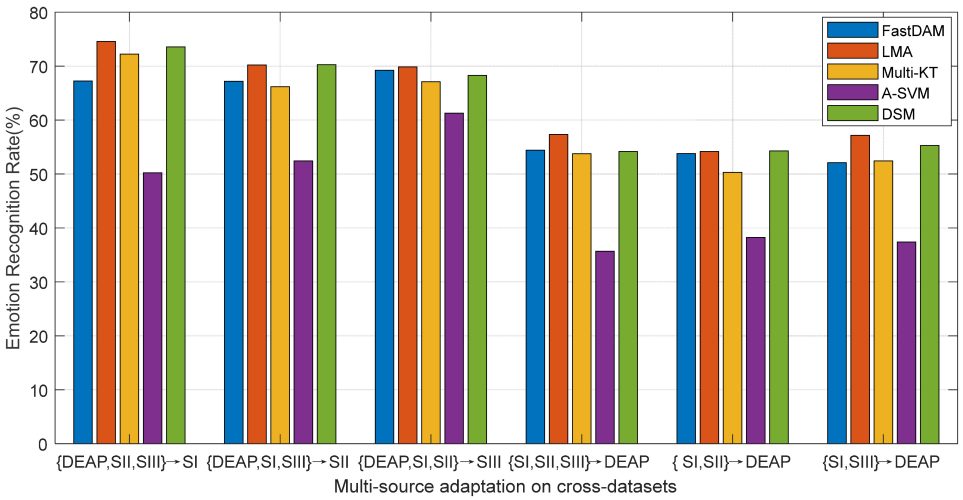


FIGURE 5 | Multi-source adaptation emotion recognition accuracy (SI: Session I, SII: Session II, SIII: Session III).

equaling to zero, we then get:

$$P_a = -\left(\Theta_a^T O_a \Theta_a + \lambda U_a\right)^{-1} M_a^T \Theta_a, \tag{36}$$

where

$$O_a = \begin{pmatrix} \vartheta_a^{q_1}\alpha^2 H_a^{-1} X_a Z_a X_a^T H_a^{-1} + \alpha^3 H_a^{-1} G_a H_a^{-1} \\ +\alpha\left(\alpha H_a^{-1} - I\right)^T \left(\alpha H_a^{-1} - I\right) \\ +\eta_a^{q_2}\alpha^2 H_a^{-1}\left((S_a^{-1} - I)G_a^{-1} D_a + S_a^{-1}\widetilde{Z}_a X_t^T\right)^T \\ \left((S_a^{-1} - I)G_a^{-1} D_a + S_a^{-1}\widetilde{Z}_a X_t^T\right) H_a^{-1} \\ +\eta_a^{q_2}\alpha^2 H_a^{-1}\left(S_a^{-1} G_v^{-1} D_v + S_a^{-1}\widetilde{Z}_a X_t^T - X_t^T\right)^T \\ \widetilde{Z}_a\left(S_a^{-1} G_v^{-1} D_v + S_a^{-1}\widetilde{Z}_a X_t^T - X_t^T\right) H_a^{-1} \\ +\beta\alpha^2 H_a^{-1} D_a^T G_a^{-1}\widetilde{\Delta} G_a^{-1} D_a H_a^{-1} \end{pmatrix}, \tag{36-1}$$

$$M_a = \begin{pmatrix} \vartheta_a^{q_1}\alpha H_a^{-1} X_a Z_a\left(\vartheta_a^q X_a^T H_a^{-1} X_a Z_a Y_a - Y\right) + \alpha^2\vartheta_a^{q_1} \\ H_a^{-1} G_a H_a^{-1} X_a Z_a Y_a + \vartheta_a^{q_1}\left(\alpha H_a^{-1} - I\right) H_a^{-1} X_a Z_a Y_a \\ +\eta_a^r\vartheta_a^{q_1}\alpha H_a^{-1}\left|(S_a^{-1} - I)G_a^{-1} D_a + S_a^{-1}\widetilde{Z}_a X_t^T\right|_F^2 H_a^{-1} \\ X_a Z_a Y_a + \beta\alpha\vartheta_a^{q_1} H_a^{-1} D_a^T G_a^{-1}\widetilde{\Delta} G_a^{-1} D_a H_a^{-1} X_a Z_a Y_a \\ +\eta_a^r\vartheta_a^{q_1}\alpha H_a^{-1}\left(S_a^{-1} G_v^{-1} D_v + S_a^{-1}\widetilde{Z}_a X_t^T - X_t^T\right)^T \widetilde{Z}_a \\ \left(S_a^{-1} G_v^{-1} D_v + S_a^{-1}\widetilde{Z}_a X_t^T - X_t^T\right) H_a^{-1} X_a Z_a Y_a \end{pmatrix}. \tag{36-2}$$

By substituting the optimal solution of the updated variables in Eqs (30), (32), (34), and (36) into Eq. (29) by mathematical calculation with the constraints $\Theta_a^T\Theta_a = I_r$, we then can get the following objective function in reference to Θ_a:

$$\Re(\Theta_a) = \min_{\Theta_a} tr\left(\Theta_a^T\left(\eta_a^{q_2} C_a - M_a\left(\Theta_a^T O_a\Theta_a + \lambda U_a\right)^{-1} M_a^T\right)\Theta_a\right), \tag{37}$$

which is equivalent to the following objective:

$$\Re(\Theta_a) = \max_{\Theta_a^T\Theta_a = I_r} tr(\Theta_a^T R_a \Theta_a), \tag{38}$$

where $R_a = M_a\left(\Theta_a^T O_a\Theta_a + \lambda U_a\right)^{-1} M_a^T - \eta_a^{q_2} C_a$. According to Li et al. (2015), Θ_a can be relaxedly obtained by the Eigen-decomposition of R_a .

Lastly, we respectively optimize ϑ_a and η_a by fixing other variables. In this situation, the objective in Eq. (29) by preserving ϑ_a changes to the following problem:

$$\min_{\vartheta_a \geq 0, \vartheta_a^T 1 = 1} \sum_{a=1}^{S} \vartheta_a^q tr(\widetilde{T}_a^T Z_a \widetilde{T}_a), \tag{39}$$

Let $g_a = tr\left(\widetilde{T}_a^T Z_a \widetilde{T}_a\right)$, the Lagrange function of Eq. (39) is

$$\Im(\vartheta_a, \varphi) = \sum_{a=1}^{S} \vartheta_a^{q_1} g_a - \varphi\left(\sum_{a=1}^{S}\vartheta_a - 1\right), \tag{40}$$

Let the derivative of $\Im(\vartheta_a, \varphi)$ with respect to ϑ_a be equivalent to 0 and we can obtain:

$$\vartheta_a = \left(\varphi / (q_1 g_a)\right)^{\frac{1}{q_1 - 1}}, \tag{41}$$

Substituting Eq. (41) into the constraint $\sum_{a=1}^{S}\vartheta_a = 1$, we obtain

$$\vartheta_a = (g_a)^{1/(1-q_1)} \Big/ \sum_{a=1}^{S}(g_a)^{1/(1-q_1)}, \tag{42}$$

With the same deduction with that of ϑ_a, we also get the following optimal solution of η_a:

$$\eta_a = (h_a)^{1/(1-q_2)} \Big/ \sum_{a=1}^{S}(h_a)^{1/(1-q_2)}, \tag{43}$$

where $h_a = tr(\Theta_a^T C_a \Theta_a) + ||F_a - F||_F^2 + tr((F_a - X_t^T T_a)^T \widetilde{Z}_a (F_a - X_t^T T_a))$.

Overall Algorithm

An overall optimization process of LMA can be outlined in the **Algorithm 1**. Following the same strategy in Zhang et al. (2019b), we employ a window-based breaking criterion to better achieve the convergence state of the algorithm. In terms of this strategy, we denote by $\hbar = 6$ the window size and compute $\varsigma = \left|MaxObj_{itr} - MinObj_{itr}\right| / MaxObj_{itr}$ in $itr-$thiteration, where $Obj_{itr} = \{Obj_{itr-\hbar+1}, ..., Obj_{itr}\}$ represents the set of historical target values in the window. While $\varsigma < \varepsilon{=}10^{-5}$, our algorithm will stop the iteration.

Algorithm 1: Multi-source adaptation learning.

Input: Source datasets $\{X_i^s\}_{i=1}^S$, Laplacian matrices $\{L_i\}_{i=1}^S$, target dataset X^t, and parameters α, β, and λ, the maximal iteration number ℓ.
Output: Converged projection matrices $\{P_i\}_{i=1}^S$, $\{\Theta_i\}_{i=1}^S$, and matrices $\{F_i\}_{i=1}^S$ and W_t.
Initialization: Set $itr = 0$, and initialize $\Theta_a = I_r$ and $\{P_a^{itr}\}_{a=1}^S$ randomly. Let $P^{itr} = [P_1^{itr}, ..., P_S^{itr}]$;
1: for $a = 1$ to S do
{
1) Compute matrix D_{itr}^a and $D_{itr}^{a(l)}$, and $\overline{K}_a^{itr}$ and $\overline{K}_{a(l)}^{itr}$ with empirical kernel mapping, thus computing
$\Lambda^a = \sum_{l=0}^c \overline{K}_a^{(l)}$, $l = 1, ..., c$ and Compute $\vartheta_a = \frac{tr\left(X_a L_a X_a^T + \Lambda^a\right)}{\sum_{a=1}^S tr\left(X_a L_a X_a^T + \Lambda^a\right)}$;
2) Initialize $T_a = \Theta_a P_a$ and $F^{itr} = \sum_{a=1}^S \vartheta_a (X^t)^T \Theta_a P_a$;
}
2: repeat
{
3) Compute W_t^{itr} by Eq. (28)
4) Compute the matrix $\widetilde{V}^{itr}$ with $(\widetilde{V})_{kk} = \frac{1}{2||(W_t^{itr})_{k,:}||_2}$;
5) set $a = 1$;
repeat
{
6) Compute the diagonal matrix Z_a^{itr}, G_a^{itr}, and $\widetilde{Z}_a^{itr}$;
7) Compute $S_a^{itr} = \widetilde{Z}_a^{itr} + I$;
8) Compute Θ_a according to Eq. (38) and then compute η_a^{itr} according to Eq. (43);
9) Compute $F_a^{itr} = (S_a^{itr})^{-1}\left(F^{itr} + \widetilde{Z}_a^{itr} X_t^T T_a^{itr}\right)$ by (30);
10) Compute F^{itr} by (32) after computing G_a^{itr} and D_a^{itr} by (33);
10) Compute T_a^{itr} by (34) with (35);
11) Compute P_a^{itr} by (36) after computing (36-1) and (36-2);
12) Compute ϑ_a^{itr} according to Eq. (42);
13) Compute the matrix
$R_a^{itr} = M_a^{itr}\left((\Theta_a^{itr})^T O_a^{itr} \Theta_a^{itr} + \lambda U_a^{itr}\right)^{-1} (M_a^{itr})^T - (\eta_a^{itr})^{q_2} C_a^{itr}$;
14) $a = a + 1$;
} **until $a > S$**
7) Update $P_a^{itr+1} = P_a^{itr}$, thus $\Theta_a^{itr+1} = \Theta_a^{itr}$ s.t. $a = 1, .., S$;
8) Update $F_a^{itr+1} = F_a^{itr}$ according to (30) s.t. $a = 1, .., S$;
9) Update ϑ_i^{itr+1} according to (42) s.t. $a = 1, .., S$;
10) Update η_i^{itr+1} according to (43) s.t. $a = 1, .., S$;
11) Update F^{itr+1} by (32), thus W_t^{itr+1} according to (28);
12) Let $itr = itr + 1$;
}**until $itr > \ell$ or $\varsigma < 10^{-5}$**
3: return $\{P_a\}_{a=1}^S$, $\{\Theta_a\}_{a=1}^S$, W_t, F and $\{F_i\}_{a=1}^S$.

In terms of the proof in Nie et al. (2010a), the convergence of the iterative procedures in **Algorithm 1** can be guaranteed by the following theorem.

Theorem 1 (Tao and Dan, 2021). The objective value in Eq. (29) would steadily decline after several iterations by **Algorithm 1**, thus finally converging to the optimum.

EXPERIMENTAL EVALUATION

In this part, we comprehensively compare the proposed method with several state of the arts on two widely used benchmark databases including SEED (Zheng and Lu, 2015) and DEAP (Koelstra et al., 2012) for EEG-based emotion recognition (Mansour et al., 2009).

Databases

According to Zhong et al. (2020) and Lan et al. (2018), there exist certain significant differences between SEED and DEAP since they can be generated by different subjects, sessions, EEG devices, experimental schemes, and emotional stimuli. Detailed information about these two datasets can be viewed in Lan et al. (2018). In the following experiments, we adopt the differential entropy (DE) (Lan et al., 2018; Zhong et al., 2020) as the data feature in emotion recognition, which has also been widely used in the preceding literatures (Shi et al., 2013; Zhang et al., 2015; Chai et al., 2016; Zheng and Lu, 2016; Chai et al., 2017; Lan et al., 2018; Zhong et al., 2020) for DA emotion recognition.

Baselines and Setting

We will systematically compare our method with such state of the arts as FSSL, an effective feature selection method without DA, FastDAM (Duan et al., 2012a), Multi-KT (Tommasi et al., 2014) with l_2-norm constraint on p, A-SVM (Yang et al., 2007), and DSM (Duan et al., 2012a). Since existing deep DA frameworks have achieved many inspiring results on emotion recognition as well as visual recognition, we also additionally present comparisons with several deep (CNN-based) DA methods with deep features: DAN (Long et al., 2015), ReverseGrad (Ganin and Lempitsky, 2015), and MultiDIAL (Carlucci et al., 2020) based on AlexNet, SDDA (Ding et al., 2018a), and CCSA (Motiian et al., 2017), a unified framework of supervised DA and generalization with deep models.

In our multi-source adaptation settings, for the baselines FSSL and A-SVM, we just equally fuse all prediction values of the base classifiers respectively obtained from each source domain[3].

In our method, LMA, there exist three vital parameters, i.e., λ, α, and β, that need to be tuned. In the community of machine learning, how to jointly search the best parameter values is still a yet unaddressed open issue. Consequently, we empirically choose these parameters using the grid search strategy also adopted in our previous work (Tao et al., 2019). Specifically, we fine-tune the values of λ, α, and β from the grid range $\{10^{-4}, 10^{-3}, ..., 10^3, 10^4\}$ in a heuristic way. Additionally, we also empirically set $q_1 = q_2 = 2$ for preventing the trivial solution in terms of the conclusion reported in Hou et al. (2017). Finally, we search the nearest neighbor number k from the set {35, 10, 15, 17}, which is also adopted in FSSL. In **Algorithm 1**, we pre-set the maximum iteration number $\tau = 100$.

Through our experiments, we adopt the RBF kernel function, i.e., $K_{i,j} = \exp(-\sigma||x_i - x_j||^2)$, for all nonlinear methods, where

[3]For each source domain, we train one SVM by using the corresponding labeled samples. Then, for each test instance x, the decision values from p SVM classifiers are converted into the probability values by using the sigmoid function (i.e., $g(t) = 1/(1 + \exp(-t))$. Finally, we average the p probability values as the final pr.

σ is equal to $1/d$. In FastDAM, we operate the same practice in Duan et al. (2012a) and set $\gamma_i = \frac{\exp(-\delta Dist(X_i^s, X))}{\sum_i \exp(-\delta Dist(X_i^s, X))}$ $(i = 1, ..., S.)$, where $\delta = 100$.

Cross-Subject Emotion Recognition

Note that different subjects even from the same dataset still have different EEG feature distributions due to the individual characteristics. We therefore practice the so-called leave-one-out cross-validation strategy conducted also in Lan et al. (2018) to evaluate the emotion recognition performance. That is, one subject remains to be the target domain, and others from the dataset are constructed as multiple sources. In this multi-source scenario, we follow the same setting as Tao and Dan (2021) to evaluate our method compared with other state of the arts on SEED and DEAP, respectively.

Performance Evaluation

We plot in **Figure 2** the recognition performance of LMA compared with the baselines on two benchmark datasets. The final obtained upper bound of chance level (UBCL) with 95% confidence interval is also recorded in **Figure 2**. It is well known that the theoretical performance (or chance level) (about 33.33 %) of the random prediction could be achieved approximately by the real chance level if the size of training data approached infinity (Lan et al., 2018). When there are finite samples, we obtain the empirical chance level by repeating the trials with the samples in question equipped with randomized class labels (Lan et al., 2018).

From **Figure 2**, we can observe that the mean performance (40.16%) of FSSL on DEAP is very close to the random prediction. While it has significantly exceeded UBCL at a 5% significance level, the relatively worse performance of FSSL still indicates the imperative importance of DA in cross-subject emotion recognition due to the substantial distribution divergence between different subjects. This importance has been witnessed by almost all baseline adaptation methods, which have yielded better performance than FSSL in all cross-subject settings. Specifically, our method, LMA, undoubtedly obtains the best recognition accuracy (about 25.14% gains over FSSL), which is closely followed by DSM. While all DA as well as our method, LMA, achieved on DEAP obvious improvement over FSSL with respect to t-test with p-value > 0.05, the mean recognition performance of these methods is yet not satisfied so far due to the complexity and difference among all subjects.

The no-adaptation method FSSL touched on SEED an average accuracy of 53.78% on three sessions from SEED, which significantly outperformed UBCL. Those multi-source adaptation methods including our method, LMA, unsurprisingly achieved more accuracy gains than the no-adaptation method on SEED. We can still observe that our method, LMA, demonstrates the best performance on SEED by upgrading the average accuracy with 75.47% w.r.t. t-test with p-value > 0.05. An interesting observation is that all methods work better on SEED than DEAP, which has also been reported in Lan et al. (2018) and Tao and Dan. (2021). The reason for this phenomenon might be that the larger distribution discrepancy between different subjects from DEAP prevented boosting performance in these methods (Mansour et al., 2009; Lan et al., 2018).

Multiple Kernel Selection

As well known, the choice of kernel is a challenging issue in the kernel learning method. Recently, multiple kernel learning (MKL) has been effectively proposed for conquering this choice issue existing in single kernel learning methods. Consequently, we also evaluate the performance boost in our method by using MKL (called as MKLMA for short) for each source domain. To this end, the first step is to construct a new space spanned by multiple kernel mapping features. We firstly denote by $\{\phi_a\}_{a=1}^{\mho}$ an empirical kernel function set, which respectively projects X_a into $\mho$ different spaces. Then, an orthogonally integrated space can be constructed by concatenating these $\mho$ spaces. We denote the mapping features in this final space by $\widetilde{\phi}(x_i) = [\phi_1(x_i)^T, \phi_2(x_i)^T, ..., \phi_{\mho}(x_i)^T]^T \in \mathbb{R}^{\mho n_a}$, where $x_i \in X_a$. Correspondingly, the kernel matrix in this final space can be easily deduced as $K_{new} = [\widetilde{K}_1; \widetilde{K}_2; ...; \widetilde{K}_{\mho}]$, where $\widetilde{K}_i$ is the ith kernel matrix from the $\mho$ feature spaces. Aiming to exploit the multiple kernel spaces, we therefore employ four kernel mapping functions including the above-used Gaussian kernel. The other additionally employed kernels are inverse square distance kernel function, Laplacian kernel function, and inverse distance kernel function, respectively, denoted as $K_{ij} = 1 \Big/ \left(1 + \sigma \left|\left|x_i - x_j\right|\right|^2\right)$, $K_{ij} = \exp\left(-\sqrt{\sigma}\left|\left|x_i - x_j\right|\right|\right)$, and $K_{ij} = 1/(1 + \sqrt{\sigma}\left|\left|x_i - x_j\right|\right|)$.

The observation from **Figure 3**, in which MKLMA significantly outperforms LMA, justifies that our LMA with MKL can further boost the recognition performance on DEAP and SEED. This also proves the importance of kernel choice in those kernel-based learning models.

Cross-Dataset Emotion Recognition

Single-Source Adaptation

In this subsection, we will demonstrate the consistent robustness of LMA by evaluating its performance in several cross-dataset settings, which is more challenging than the cross-subject adaptation due to the intrinsic difference between datasets. For the scenario of cross-dataset adaptation, we specially design several different cross-dataset strategies by splitting the training set and test set, respectively, in terms of their EEG instruments and emotional stimuli sources, thus making up six cases, i.e., *DEAP* $\rightarrow$ *SI*, *DEAP* $\rightarrow$ *SII*, *DEAP* $\rightarrow$ *SIII*, *SI* $\rightarrow$ *DEAP*, *SEED II* $\rightarrow$ *DEAP*, and *SIII* $\rightarrow$ *DEAP*, where A$\rightarrow$B denotes the adaptation from the dataset A to the dataset B, and SI, SII, and SIII are respectively denoted as the dataset of session I, session II, and session III from the database SEED.

A representative hypothesis used in DA is that the feature space of both source and target domains should be the same. Following this assumption, we employ in this part only 32 channels shared between SEED and DEAP to construct a common feature space with 160 dimensions for both domain datasets. In the first three trials, we sample 2,520 samples as the source from DEAP and 2,775 samples as the target from three different sessions (SI, SII, and SIII) in SEED. We evaluate each subject with respect to recognition accuracy in each session and

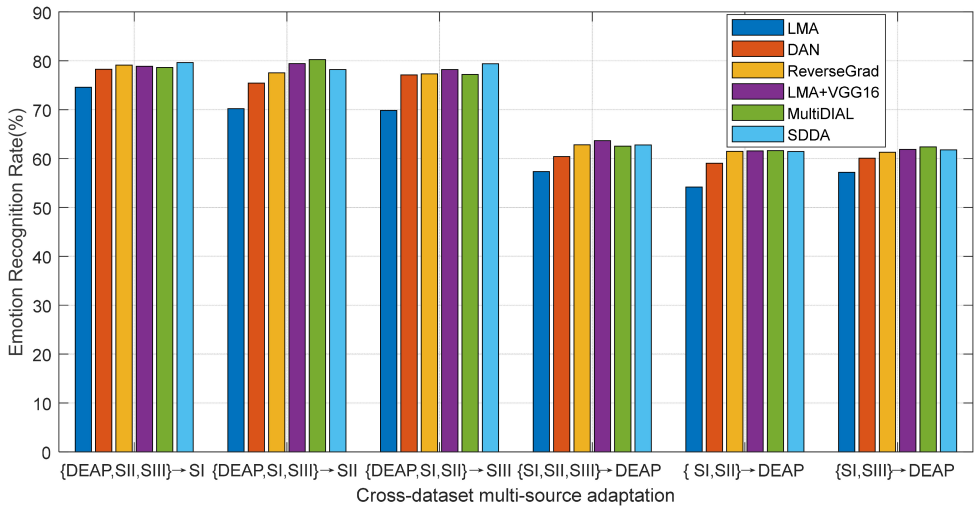


FIGURE 6 | Emotion recognition accuracies of different methods using deeply extracted features (SI: Session I, SII: Session II, SIII: Session III).

TABLE 2 | Multi-source adaptation emotion recognition accuracies of derived methods as well as LMA.

Method	{DEAP,SII,SIII} →SI	{DEAP,SI,SIII} →SII	{DEAP,SI,SII} →SIII	{SI,SII,SIII} →DEAP	{SI,SII} →DEAP	{SI,SIII} →DEAP
LMA_NF	73.52	69.10	69.58	55.43	52.01	56.22
LMA_NL	69.13	65.36	66.11	52.32	53.16	52.71
LMA_NS	72.68	68.23	68.38	54.69	51.08	54.20
LMA	**74.57**	**70.2**	**69.85**	**57.33**	**54.16**	**57.17**

Bold denotes the best recognition rates (SI: Session I, SII: Session II, SIII: Session III).

then record the final mean results over 15 subjects from SEED. In the other trials, we resample 41,625 source samples from SEED and 180 target samples from DEAP. We also record the mean recognition accuracy of each subject in DEAP over 14 subjects. For the limitation of memory, 10% of the source data (4,162 samples) is randomly sampled as the actual training samples (Shi et al., 2013; Zheng and Lu, 2015, 2016; Chai et al., 2016, 2017; Lan et al., 2018; Zhong et al., 2020).

The mean recognition results on six cross-datasets are plotted in **Figure 4**. We can observe from these results that the performance of FSSL is almost near the random guess in that it is slightly inferior to UBCL with about 95% confidence interval. Besides, as observed from the results, the mean performance of each method is slightly worse on cross-dataset than within-dataset. This confirms the larger distribution gaps between two datasets than within-dataset. The advantage of DA would be reflected in this situation since DA could potentially relieve the distribution issue in the cross-dataset applications, which can also be justified by the observation from **Figure 4**, where all DA methods outperform the no-adaptation one. While Multi-KT and FastDAM occasionally obtain the best performance in some settings, our method, LMA, still contributes the best performance in most cases.

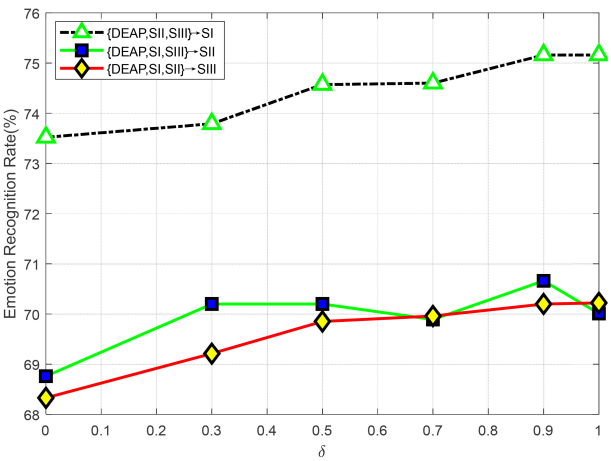


FIGURE 7 | Emotion recognition accuracies with different values of δ (SI: Session I, SII: Session II, SIII: Session III).

Multi-Source Adaptation

As reported in preceding works about DA learning, multiple source domains can improve the adaptation performance to some extent by integrating multiple prior knowledge. Nevertheless, in concrete applications, multi-source adaptation also incurs another challenge, i.e., source scalability issue, since multi-source learning could lead to the so-called "negative transfer" problem. In this scenario, how to discriminately exploit multiple sources becomes a challenge worthy to be addressed in the

multi-source adaptation learning frameworks. To this end, we will explore in this part the different reliabilities of the prior sources in the emotion recognition task (Tao et al., 2019; Tao and Dan, 2021). We evaluate the performance of all baseline DA methods with multiple prior sources on the designed cross-dataset settings. The average accuracies of all methods are plotted in **Figure 5**, where A-SVM employs the average prior model.

When there exists very large distribution discrepancy between different domain datasets, it is hard for A-SVM to eliminate the inter-domain distribution bias. Therefore, the results in **Figure 5** show that A-SVM is inferior to other multi-source adaptation methods in most settings. A-SVM even has a downgraded performance tendency with the increase of source domains in some scenarios, which indicates the existence of a "negative transfer" phenomenon in A-SVM. Another interesting observation from **Figure 5** is that all DA methods except A-SVM achieve more improvement by leveraging multiple source knowledge than by bridging only one source (i.e., cross-subject settings) when the number of source domains increase. This proves that it is beneficial to leverage multiple sources for boosting the recognition performance. Moreover, LMA and DSM conquer others by touching on the top performance, due to their designed weights for discriminately screening the optimal sources. Our method, LMA, obtains more gains over DSM in some scenarios. A possible explanation is that the shared discriminative information among source models in LMA is advantageous to multi-source adaptation learning by utilizing the optimal weight vector.

Deep Feature Adaptation

In this subsection, we will particularly evaluate our method, LMA, with deeply extracted features by comparing it with several recently proposed deep adaptation models on cross-dataset emotion recognition using the multi-source settings.

In practical tasks, our method, LMA, can be trained on the deeply transformed features of all domains, which follows the same setup with that in Zhou et al. (2018) and Zhu et al. (2017). Concretely, some pretrained deep model (e.g., VGG16) is first fine-tuned using the source domain, then the deep features can be extracted from EEG signals in both source and target domains with this CNN model, and finally the recognition model would be trained on these extracted features. In the context of our experiments, we denote our methods with the VGG16 model as LMA+VGG16. As for DAN, SDDA, MultiDIAL, and ReverseGrad, we use their released source codes to fine-tune the pre-trained models by respectively using the pre-tuned parameters in their works (Ganin and Lempitsky, 2015; Long et al., 2015). Note that these deep adaptation methods typically aim to learn domain-invariant representations. Different from the deep adaptation frameworks, our proposed method explores to learn a domain-invariant recognition model with strong generalization ability from the source domain to the target domain. Consequently, we expect that our method can further upgrade the recognition performance with the co-learning strategy on the deeply extracted features from some deep model.

We plot the mean results of all methods in **Figure 6**, from which we can observe that our deep adaptation method LMA+VGG16 significantly outperforms LMA. This indicates the advantage of deep features, which can be attributed to its robust feature representation. Furthermore, LMA+VGG16 also obtains comparable recognition performance with respect to other deep adaptation methods. This may be attributed to the classification-level constraint in LMA, where most of the source discriminative structures are expected to be preserved by the guidance of target classification. In some cases, as shown in **Figure 6**, LMA+VGG16 even achieves the top performance compared with other deep adaptation frameworks. This phenomenon shows that the proposed LMA can become an effective surrogate to the deep adaptation model by just exploiting the deep features extracted from any one of the state-of-the-art deep models.

Parameter Impact

In our method, LMA, there exist three hyper-parameters (i.e., λ, β, and α) that needed to be tuned. These hyper-parameters are mainly used to trade off different components of the LMA framework. We therefore respectively set these parameters into their extreme values to explore the importance of each component in LMA. To this end, we set $\beta = 0$ to denote LMA without target feature selection by LMA_NF, set $\alpha = \eta_a = 0$ to denote by LMA_NL the case where LMA ignores latent space representations, and set $\lambda = 0$ to denote by LMA_NS the scenario where LMA fails to consider the shared discriminative structures among multiple sources. We evaluate these derived methods on cross-dataset recognition tasks.

The results in **Table 2** clearly show that none of the three derived methods can achieve the best performance as that obtained by LMA. This further verifies the valuable contribution of each component to LMA. Specifically, LMA_NL has a significant downgraded manifestation compared with LMA, which, from the opposite side, proves that the utilization of shared latent spaces is very preferable to boosting the performance of LMA; the performance of LMA_NTF is slightly weaker than LMA, i.e., the performance of LMA would be slightly impacted by the target feature selection due to the intrinsic existence of some noise/outlier data in the target data; the inferior performance of LMA_NS proves the importance of the utilization of correlation knowledge among source models in cross-dataset emotion recognition.

Note that in section Emotion Recognition, we use the following combination function to recognize the emotion level of the given test data x_i^t:

$$j = \arg\max_j \left(y_i^t = \delta f^s(x_i^t) + (1-\delta) f^t(x_i^t)\right)_j,$$

where $\delta \in [0, 1]$ is a trade-off parameter, which is empirically set as 0.5 in the preceding trials. In this part, we will further evaluate the impact on LMA with different values of δ in multi-source adaptation scenarios. We plot the recognition accuracy w.r.t. different values of δ in **Figure 7**. From the curves shown in **Figure 7**, we can observe the following several interesting results:

(1) Theoretically, δ controls the weight of source classifiers and larger values of δ will make the source classifiers more important in LMA. An extreme case is $\delta \to 1$, where only source classifiers are guaranteed, but the target discriminative information for the test samples is discarded. In this case, all experimental results demonstrate a trend of slight downgrade. This shows the necessity of composite discrimination information by combining both source and target classifiers.
(2) Another extreme case is $\delta \to 0$. In this case, LMA will recognize the emotion state of certain test data by only using the target discriminator, which cannot leverage the prior source information with discriminating power. From **Figure 7**, we can see that all curves show an obvious upgrade in performance around $\delta = 0$, which shows the importance of multi-source discriminative models in our framework.
(3) We cannot obtain the best performance when δ values are relatively small or large, which shows the significance of exploiting the discriminative information from both source and target classifiers in our method.
(4) After $\delta > 0.5$, we can see that most curves are relatively stable across δ values, which shows that our method is not significantly sensitive to $\delta > 0.5$. Hence, we can empirically set $\delta = 0.5$ in the experiments.

CONCLUSION

To deal with the cross-subject/dataset EEG-based emotion recognition task, we proposed a robust LMA. In multiple domain-invariant latent spaces, LMA aims at transferring multi-source knowledge into target learning mainly by leveraging correlation knowledge among source models, which discriminatively screens unimportant prior evidences in sources. The comprehensive experiments performed on two public datasets verify the effectiveness of LMA in dealing with cross-subject/dataset emotion recognition. In most scenarios, our LMA (or LMA-VGG16) obtains the best results or comparable performance with respect to several representative baselines.

Since the implementation of LMA algorithm needs an iterative optimization procedure, how to improve the efficiency of LMA and seek a more efficient algorithm would be an issue worthy of further study in our future research. The unreliable and misleading pseudo labels strategy may be another potential problem in our LMA. Consequently, our successive work would explore how to seamlessly incorporate target label into the framework of LMA.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by the research institution special program of Ningbo Polytechnic under Grant No. NZ21JG006, and Zhejiang Provincial Natural Science Foundation of China under Grant No. Lgg20F020013.

REFERENCES

Bao, G., Zhuang, N., Tong, L., Yan, B., Shu, J., Wang, L., et al. (2021). Two-level domain adaptation neural network for EEG-based emotion recognition. *Front. Hum. Neurosci.* 14:605246. doi: 10.3389/fnhum.2020.605246

Bruzzone, L., and Marconcini, M. (2010). Domain adaptation problems: A DASVM classification technique and a circular validation strategy. *IEEE Trans.* 32, 770–787. doi: 10.1109/TPAMI.2009.57

Carlucci, F. M., Porzi, L., Caputo, B., Ricci, E., and Buló, S. R. (2020). MultiDIAL: domain alignment layers for (Multisource) unsupervised domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*43, 4441–4452. doi: 10.1109/TPAMI.2020.3001338

Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., et al. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* 17:1014. doi: 10.3390/s17051014

Chai, X., Wang, Q., Zhao, Y., Liu, X., Bai, O., and Li, Y. (2016). Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. *Comput. Biol. Med.* 79, 205–214. doi: 10.1016/j.compbiomed.2016.10.019

Chen, B., Lam, W., Tsang, I. W., and Wong, T. L. (2013). Discovering low-rank shared concept space for adapting text mining models. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1284–1297. doi: 10.1109/TPAMI.2012.243

Chu, W. S., De la Torre, F., and Cohn, J. F. (2017). Selective transfer machine for personalized facial action unit detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 529–545. doi: 10.1109/CVPR.2013.451

Ding, Z., Nasrabadi, N. M., and Fu, Y. (2018a). Semi-supervised deep domain adaptation via coupled neural networks. *IEEE Trans. Image Process.* 27, 5214–5224. doi: 10.1109/TIP.2018.2851067

Ding, Z., Nasrabadi, N. M., and Fu, Y. (2018b). Incomplete multisource transfer learning. *IEEE Trans. Neural Networks Learn. Syst.* 29, 310–323. doi: 10.1109/TNNLS.2016.2618765

Ding, Z., Sheng, L., Ming, S., and Fu, Y. (2018c). "Graph adaptive knowledge transfer for unsupervised domain adaptation," in *Proceedings of the 15th European Conference (ECCV2018)* (Munich).

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194.

Duan, L., Tsang, I. W., and Xu, D. (2012a). Domain transfer multiple kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 465–479. doi: 10.1109/TPAMI.2011.114

Duan, L., Xu, D., and Fu, C. S. (2012b). "Exploiting web images for event recognition in consumer videos: a multiple source domain adaptation approach," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition* (Providence, RI), 1338–1345.

Duan, L., Xu, D., and Tsang, I. W. (2012c). Domain adaptation from multiple sources: a domain-dependent regularization approach. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 504–518. doi: 10.1109/TNNLS.2011.2178556

Ganin, Y., and Lempitsky, V. (2015). "Unsupervised domain adaptation by back-propagation," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning (ICMR)*, Lille, 1180–1189.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., et al. (2016). Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* 17, 1–35. doi: 10.1109/TNNLS.2020.3025954

Ghifary, M., Balduzzi, D., Kleijn, W. B., and Zhang, M. (2017). Scatter component analysis: a unified framework for domain adaptation and domain generalization. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1414–1430. doi: 10.1109/TPAMI.2016.2599532

Gretton, A., Fukumizu, K., Harchaoui, Z., and Sriperumbudur, B. K. (2009). "A fast, consistent kernel two-sample test," in *Proceedings of the Conference on Neural Information Processing Systems 22*, eds Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Vancouver, BC: MIT Press), 673–681.

Hou, C., Nie, F., Tao, H., and Yi, D. (2017). Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Trans. Knowl. Data Eng.* 29, 1998–2011. doi: 10.1109/tkde.2017.2681670

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Magaz.* 11, 20–31. doi: 10.1109/mci.2015.2501545

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/taffc.2014.2339834

Kim, M.-K., Kim, M., Oh, E., and Kim, S.-P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Comput. Math. Methods Med.* 2013:573734. doi: 10.1155/2013/573734

Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/t-affc.2011.15

Lan, Z., Sourina, O., Wang, L., Scherer, R., and Müller-Putz, G. R. (2018). Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. *IEEE Trans. Cogn. Dev. Syst.* 11, 85–94. doi: 10.1109/tcds.2018.2826840

Li, H., Wan, R., Wang, S., and Kot, A. C. (2021). Unsupervised domain adaptation in the wild via disentangling representation learning. *Int. J. Comput. Vis.*129, 267–283. doi: 10.1007/s11263-020-01364-5

Li, J., Qiu, S., Du, C., Wang, Y., and He, H. (2020a). Domain adaptation for EEG emotion recognition based on latent representation similarity. *IEEE Trans. Cogn. Dev. Syst.* 12, 344–353. doi: 10.1109/tcds.2019.2949306

Li, J., Qiu, S., Shen, Y.-Y., Liu, C.-L., and He, H. (2020b). Multisource transfer learning for cross-subject EEG emotion recognition. *IEEE Trans. Cyber.* 50, 3281–3293. doi: 10.1109/TCYB.2019.2904052

Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., and Hu, B. (2018a). Exploring EEG features in cross-subject emotion recognition. *Front. Neurosci.* 12:162. doi: 10.3389/fnins.2018.00162

Li, Y., Zheng, W., Cui, Z., Zhang, T., and Zong, Y. (2018b). "A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)* (Palo Alto, CA).

Li, Y., Zheng, W., Cui, Z., Zong, Y., and Ge, S. (2018c). EEG emotion recognition based on graph regularized sparse linear regression. *Neural Process. Lett.* 49, 1–17. doi: 10.1109/taffc.2020.2994159

Li, Z., Liu, J., Tang, J., and Lu, H. (2015). Robust structured subspace learning for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 2085–2098.

Liu, Q., and Liu, H. (2021). Criminal psychological emotion recognition based on deep learning and EEG signals. *Neural Comput. Appl.* 33, 433–447. doi: 10.1007/s00521-020-05024-0

Long, M., Cao, Y., Wang, J., and Jordan, M. (2015). "Learning transferable features with deep adaptation networks," in *Proceedings of the 32nd International Conference on Machine Learning (PMLR)*, Lille, 97–105.

Long, M., Wang, J., Ding, G., Pan, S. J., and Yu, P. S. (2014). Adaptation Regularization: A General Framework for Transfer Learning. *IEEE Trans. Knowl. Data Eng.* 26, 1076–1089. doi: 10.1109/tkde.2013.111

Lotfi, E., and Akbarzadeh-T, M.-R. (2014). Practical emotional neural networks. *Neural Netw.* 59, 61–72. doi: 10.1016/j.neunet.2014.06.012

Mansour, Y., Mohri, M., and Rostamizadeh, A. (2009). "Domain Adaptation with Multiple Sources," in *Proceedings of the Conference on Neural Information Processing Systems* (Vancouver, BC: MIT Press), 1041–1048.

Motiian, S., Piccirilli, M., Adjeroh, D. A., and Doretto, G. (2017). "Unified deep supervised domain adaptation and generalization," in *Proceedings of the IEEEInternational Conference on Computer Vision ICCV* (Venice).

Mühl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain Comput. Interfaces* 1, 66–84. doi: 10.1080/2326263x.2014.912881

Musha, T., Terasaki, Y., Haque, H. A., and Ivamitsky, G. A. (1997). Feature extraction from EEGs associated with emotions. *Artif. Life Robot.* 1, 15–19. doi: 10.1007/bf02471106

Nie, F., Huang, H., Cai, X., and Ding, C. (2010a). "Efficient and robust feature selection via Joint l2,1-norms minimization," in *Proceedings of the 22th Annual Conference Neural Information Processing Systems* (Cambridge, MA: MIT Press), 1813–1821.

Nie, F., Xu, D., Tsang, I. W., and Zhang, C. (2010b). Flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction. *IEEE Trans. Image Process.* 19, 1921–1932. doi: 10.1109/TIP.2010.2044958

Pan, S. J., Tsang, I. W., Kwok, J. T., and Yang, Q. (2011). Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* 22, 199–210. doi: 10.1109/TNN.2010.2091281

Pandey, P., and Seeja, K. (2019). "Emotional state recognition with EEG signals using subject independent approach," in *Data Science and Big Data Analytics. Lecture Notes on Data Engineering and Communications Technologies*, Vol. 16, eds D. Mishra, X. S. Yang, and A. Unal (Singapore: Springer), 117–124. doi: 10.1007/978-981-10-7641-1_10

Rosenstein, M. T., Marx, Z., and Kaelbling, L. P. (2005). "To transfer or not to transfer," in *Proceedings of the Conference on Neural Information Processing Systems*, (Cambridge, MA: MIT Press).

Shi, L. C., Jiao, Y. Y., and Lu, B. L. (2013). "Differential entropy feature for EEG- based vigilance estimation. 2013," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Osaka), 6627–6630.

Shi, X., Guo, Z., Lai, Z., Yang, Z. Bao, and Zhang, D. (2015). A framework of joint graph embedding and sparse regression for dimensionality reduction. *IEEE Trans. Image Process.* 24, 1341–1355. doi: 10.1109/TIP.2015.2405474

Song, T., Zheng, W., Song, P., and Cui, Z. (2018). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* 11, 532–541. doi: 10.3389/fnbot.2022.834952

Tao, J., Chung, F. L., and Wang, S. (2012). On minimum distribution discrepancy support vector machine for domain adaptation. *Pattern Recogn.* 45, 3962–3984. doi: 10.1016/j.patcog.2012.04.014

Tao, J., and Dan, Y. (2021). Multi-source Co-adaptation for EEG-based emotion recognition by mining correlation information. *Front. Neurosci.* 15:677106. doi: 10.3389/fnins.2021.677106

Tao, J., Wen, S., and Hu, W. (2015). L1-norm locally linear representation regularization multi-source adaptation learning. *Neural Netw.* 69, 80–98. doi: 10.1016/j.neunet.2015.01.009

Tao, J., Wen, S., and Hu, W. (2016). Multi-source adaptation learning with global and local regularization by exploiting joint kernel sparse representation. *Knowl. Based Syst.* 98, 76–94. doi: 10.1016/j.knosys.2016.01.021

Tao, J., and Xu, H. (2019). Discovering domain-invariant subspace for depression recognition by jointly exploiting appearance and dynamics feature representations. *IEEE Access* 99, 186417–186436. doi: 10.1109/access.2019.2961741

Tao, J., Zhou, D., Liu, F., and Zhu, B. (2019). . Latent multi-feature co-regression for visual recognition by discriminatively leveraging multi-source models. *Pattern Recogn.* 87, 296–316. doi: 10.1016/j.neunet.2019.02.007

Tao, J. W., Song, D., Wen, S., and Hu, W. (2017). Robust multi-source adaptation visual classification using supervised low-rank representation. *Pattern Recogn.* 61, 47–65. doi: 10.1016/j.patcog.2016.07.006

Tommasi, T., Orabona, F., and Caputo, B. (2014). Learning categories from few examples with multi-model knowledge transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 928–941. doi: 10.1109/TPAMI.2013.197

Tzeng, E., Hoffman, J., Darrell, T., and Saenko, K. (2015). "Simultaneous deep transfer across domains and tasks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, 4068–4076. doi: 10.1109/ICCV.2015.463

Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017). "Adversarial Discriminative Domain Adaptation," in *Proceedings of the 2017 IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), (Honolulu, HI), 2962–2971.

Wang, F., Zhang, W., Xu, Z., Ping, J., and Chu, H. (2021). A deep multi-source adaptation transfer network for cross-subject electroencephalogram emotion recognition. *Neural Comput. Appl.* 33, 9061–9073. doi: 10.1007/s00521-020-05670-4

Yan, S., Xu, D., Zhang, B., Zhang, H. J., Yang, Q., and Lin, S. (2006). Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* 29:40. doi: 10.1109/TPAMI.2007.12

Yang, J., Yan, R., and Hauptmann, A. G. (2007). "Cross-domain video concept detection using adaptive svms," in *Proceedings of the 15th ACM International Conference on Multimedia. ACM*, (Augsburg), 188–197.

Yang, Y., Ma, Z., Hauptmann, A. G., and Sebe, N. (2013). Feature selection for multimedia analysis by sharing information among multiple tasks. *IEEE Trans. Multimedia* 15, 661–669. doi: 10.1109/tmm.2012.2237023

Zhang, K., Gong, M., and Schölkopf, B. (2015). "Multi-source domain adaptation: A causal view," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, (Menlo Park, CA), 3150–3157.

Zhang, Y., Chung, F., and Wang, S. (2020a). Clustering by transmission learning from data density to label manifold with statistical diffusion. *Knowl. Based Syst.* 193, 105330. doi: 10.1016/j.knosys.2019.105330

Zhang, Y., Chung, F. L., and Wang, S. (2019a). Takagi-sugeno-kang fuzzy systems with dynamic rule weights. *J. Intell. Fuzzy Syst.* 37, 8535–8550. doi: 10.1016/j.isatra.2017.10.012

Zhang, Y., Dong, J., Zhu, J., and Wu, C. (2019b). Common and special knowledge-driven TSK fuzzy system and its modeling and application for epileptic EEG signals recognition. *IEEE Access* 7, 127600–127614. doi: 10.1109/access.2019.2937657

Zhang, Y., Li, J., Zhou, X., Zhou, T., Zhang, M., Ren, J., et al. (2019c). A view-reduction based multi-view TSK fuzzy system and its application for textile color classification. *J. Ambient Intell. Hum. Comput.* 9, 1–11. doi: 10.1007/s12652-019-01495-9

Zhang, Y., Tian, F., Wu, H., Geng, X., Qian, D., Dong, J., et al. (2017). Brain MRI tissue classification based fuzzy clustering with competitive learning. *J. Med. Imaging Health Inform.* 7, 1654–1659. doi: 10.1006/cbmr.1996.0023

Zhang, Y., Wang, L., Wu, H., Geng, X., Yao, D., and Dong, J. (2016). A clustering method based on fast exemplar finding and its application on brain magnetic resonance images segmentation. *J. Med. Imaging Health Inform.* 6, 1337–1344. doi: 10.1166/jmihi.2016.1923

Zhang, Y., Wang, S., Xia, K., Jiang, Y., and Qian, P. (2020b). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inform. Fusion* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002

Zheng, W. (2017). Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. *IEEE Trans. Cogn. Dev. Syst.* 9, 281–290. doi: 10.1109/tcds.2016.2587290

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Autonom. Ment. Dev.* 7, 162–175. doi: 10.1109/tamd.2015.2431497

Zheng, W. L., and Lu, B. L. (2016). "Personalizing EEG-based affective models with transfer learning," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, (New York, NY), 2732–2738.

Zheng, W. L., Zhang, Y. Q., Zhu, J. Y., and Lu, B. L. (2015). "Transfer components between subjects for EEG-based emotion recognition," in *Proceedings of the2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, (Xi'an), 917–922.

Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* 99:1.

Zhou, X., Jin, K., Shang, Y., and Guo, G. (2018). visually interpretable representation learning for depression recognition from facial images. *IEEE Trans. Affect. Comput.* 11, 542–552. doi: 10.1109/TAFFC.2018.2828819

Zhu, Y., Shang, Y., Shao, Z., and Guo, G. (2017). Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Trans. Affect. Comput.* 9, 578–584. doi: 10.1109/TAFFC.2017.2650899

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

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 04 May 2022
doi: 10.3389/fnins.2022.855421

Multi-Model Adaptation Learning With Possibilistic Clustering Assumption for EEG-Based Emotion Recognition

Yufang Dan[1,2*†], *Jianwen Tao*[1†] *and Di Zhou*[3*]

[1] *Institute of Artificial Intelligence Application, Ningbo Polytechnic, Ningbo, China,* [2] *Key Laboratory of 3D Printing Equipment and Manufacturing in Colleges and Universities of Fujian Province, Fujian, China,* [3] *Industrial Technological Institute of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou, China*

In machine learning community, graph-based semi-supervised learning (GSSL) approaches have attracted more extensive research due to their elegant mathematical formulation and good performance. However, one of the reasons affecting the performance of the GSSL method is that the training data and test data need to be independently identically distributed (IID); any individual user may show a completely different encephalogram (EEG) data in the same situation. The EEG data may be non-IID. In addition, noise/outlier sensitiveness still exist in GSSL approaches. To these ends, we propose in this paper a novel clustering method based on structure risk minimization model, called multi-model adaptation learning with possibilistic clustering assumption for EEG-based emotion recognition (MA-PCA). It can effectively minimize the influence from the noise/outlier samples based on different EEG-based data distribution in some reproduced kernel Hilbert space. Our main ideas are as follows: (1) reducing the negative impact of noise/outlier patterns through fuzzy entropy regularization, (2) considering the training data and test data are IID and non-IID to obtain a better performance by multi-model adaptation learning, and (3) the algorithm implementation and convergence theorem are also given. A large number of experiments and deep analysis on real DEAP datasets and SEED datasets was carried out. The results show that the MA-PCA method has superior or comparable robustness and generalization performance to EEG-based emotion recognition.

Keywords: semi-supervised learning, multi-model adaptation, clustering assumption, encephalogram, fuzzy entropy, emotion recognition

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Liang Yu,
Shanghai Jiao Tong University, China
Tingyang Chen,
Wuhan University of Technology, China

***Correspondence:**
Yufang Dan
2018020@nbpt.edu.cn
Di Zhou
sion2005@sasu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *15 January 2022*
Accepted: *25 February 2022*
Published: *04 May 2022*

Citation:
Dan Y, Tao J and Zhou D (2022) Multi-Model Adaptation Learning With Possibilistic Clustering Assumption for EEG-Based Emotion Recognition. Front. Neurosci. 16:855421. doi: 10.3389/fnins.2022.855421

INTRODUCTION

Emotion is a psychological experience from human beings of the world, which is complex and changeable (Dolan, 2002; Zhang et al., 2016, 2019b). Different human beings have different emotional understanding on the same thing and may make misjudgment about the emotion occasionally, let alone machines. Therefore, emotion recognition has attracted great attention from researchers (Kim et al., 2013; Mühl et al., 2014; Zhao et al., 2015, 2016; Chu et al., 2017). In this paper, we mainly recognize the corresponding emotion by the internal changes of the human body which include the heart rate, blood pressure, respiratory rate,

magneto encephalogram, electroencephalogram (Mühl et al., 2014), and so on. Generally, most existing EEG-based emotion recognition systems are divided into two steps: data preparation and classifier training (Lan et al., 2018; Zhang et al., 2020). EEG feature extraction methods are comprehensively sorted out in Jenke et al. (2014). In order to improve the recognition accuracy, there existed many EEG-based emotion recognition approaches (Musha et al., 1997; Kim et al., 2013). A satisfactory state of emotion detection based on brain computer interface (BCI) is to detect the emotional state by a real-time EEG signal without inputting signal from the subjects (Zhang et al., 2019b). Different feedback is given to different emotional states in the meantime. These proposed methods (Zhang et al., 2016, 2017) are used to recognize multiple emotion classes from EEG. The latest affective BCIs adopted machine learning algorithms and relied on a few support vectors (Jenke et al., 2014; Mühl et al., 2014). For recording of the EEG signal of the expected target emotion, it is necessary to provide emotional stimulation of the expected concrete emotion to subjects. In the training/calibration stage, the EEG datasets with labels are used to train the emotion recognizer. Many researchers have reported sound classification performance on emotion recognition from real-time EEG data (Mühl et al., 2014).

Due to the high cost of obtaining labeled data, semi-supervised learning (SSL) technology has appeared. It only needs a small, labeled data and a large, unlabeled data to learn a model, which solves the problem on supervised learning needing a large number of labeled samples. Tu and Sun (2013) presented an EEG classifier *via* SSL feature extraction strategy. Tao et al. (2015, 2016, 2017) and Wu and Deng (2018) showed a SSL method for reducing the possible negative impact from random initialization parameters on neural networks. Zu et al. (2019) proposed that remote sensing image classification method based on SSL can effectively improve the accuracy of land cover classification and has a higher efficiency in remote sensing image classification since graph-based semi-supervised learning (GSSL) (Li and Zhou, 2011; Liu et al., 2012; Wang et al., 2012), with its good performance, has been extensively studied. The manifold regularization (MR) (Belkin et al., 2006; Gao et al., 2010; Nie et al., 2010) is a popular GSSL method. A general MR framework was presented by Nie et al. (2010).

In general, the clustering assumption is a basic assumption in GSSL: similar samples should belong to the same class (Chapelle, 2006; Zhu and Goldberg, 2009; Xue et al., 2011; Zhou et al., 2014; Wang et al., 2019). In other words, each sample only have one label, which we called hard classification. However, in the real applications of emotion recognition, its performance will be discounted by this assumption—for example, for different subjects in different scenes, crying may be understood as sad and happy.

To handle the limitation from this assumption, Wang et al. (2012) and Zhang et al. (2019b) proposed a novel clustering assumption that can significantly boost the classifier performance. Assuming that similar samples have the same label membership, each sample may have multiple membership values (Zhang et al., 2019a), not only one. However, this method has a constraint in that the sum of the membership values of each sample is 1. This constraint may lead to the membership values of some noise being close to or even greater than those of normal samples; it may lead to misrecognition.

According to this problem in the SSCCM method, Dan et al. (2021) proposed SSPCA that relaxed the constraint in SSCCM and added a fuzzy entropy regularization term (Kosko, 1986; Krishnapuram and Keller, 1993; Zhang et al., 2019c) for increasing the samples' discriminative information to get a membership function with better generalization and that can also reduce the negative impact of noise and outlier on recognition performance to improve the robustness of the method. Wang and Chen (2013) designed SA-SSCCM. Specifically, the SSCCM method is upper boundary and the LS-SVM method is lower boundary, respectively—that is, if unlabeled data is good for model training, the classification result of SA-SSCCM is close to SSCCM; if unlabeled data penalizes model training, the classification result of SA-SSCCM is close to that of the LS-SVM method, and the interference of noise data to the SA-SSCCM model training is avoided. However, both SSPCA method and SA-SSCCM method require training data, and the test data should be independently identically distributed (IID). Due to the difference among different subjects in real emotion recognition applications, it may cut recognition accuracy. It is hard to guarantee that two datasets are IID.

Toward the problem of reduced recognition accuracy caused by non-IID of training data and test data, this paper adopts domain adaptation learning (DAL) (Bruzzone and Marconcini, 2010; Tao et al., 2021) related to computer vision and machine learning (Bishop, 2006; Zhu, 2008). Generally, DAL includes instance-based DAL, feature-based DAL, and model-based DAL (Pan and Yang, 2010). The instance-based DAL and feature-based DAL need to access instances from the source domain during the model learning. When the source dataset is relatively large, the training efficiency will be reduced. The model-based DAL uses the pre-trained source classifier on some source datasets to learn an effective target classifier, which has good classification effectiveness and high efficiency. Much more DAL categories in Pan and Yang (2010) can be found. Therefore, this paper proposes a multi-model adaptation learning with possibilistic clustering assumption for EEG-based emotion recognition (MA-PCA).

The main ideas are as follows: firstly, according to manifold learning (Tenenbaum et al., 2000; Belkin and Niyogi, 2001; Gao et al., 2010; Nie et al., 2010), there is similarity among samples within its local. According to formula (1) in Dan et al. (2021), the local weighted mean (LWM) point is determined by the convex hull of the k-nearest neighbors. It represents the mean value of the local. Therefore, the neighbors in the local should have consistency with the mean value—that is, the labels of each neighbor in the local and its corresponding LWM should be similar (or consistent). Then, it is assumed that each neighbor has a similar label membership to its corresponding LWM (Bottou and Vapnik, 1992; Atkeson et al., 1997; Xue and Chen, 2007); secondly, the classification prediction results are mutually verified by the decision function and the membership function to improve the classification reliability; thirdly, a fuzzy entropy regularization term is proposed to increase the sample discrimination information; then, we can get a membership function with better generalization, and the negative impact of noise and outlier will be relaxed on

recognition performance. Finally, a classification model with better generalization performance is obtained by adding a multi-model adaptation regularization term for IID and non-IID on training data and test data, respectively. The major contributions of this work are the following:

(1) A multi-model adaptation learning with possibilistic clustering assumption for EEG-based (MA-PCA) is proposed.

(2) Since a multiple auxiliary discriminant model is good for SSL with a small labeled instance, the regular term has Laplacian local consistency and the regularize term has different weights of multiple-source models. It aims to expand the discriminant space of the target domain and guarantee the local structure consistency among inner samples of source domain and target domain. At the end, it solves the non-IID problem between the training data and test data, too.

(3) Finally, our comprehensive experiments on real datasets (i.e., DEAP, SEED) show that the method has better robustness and generalization.

The remainder of this work is organized as follows: In section "Proposed Framework," our framework MA-PCA will be designed which includes MA-PCA formulation, optimization, and convergence analysis, and section "Algorithm of MA-PCA" arranges the corresponding optimal algorithm of MA-PCA. The experimental results and analysis on two real EEG datasets (i.e., DEAP and SEED) are presented in section "Experimental Evaluation." Finally, we conclude in section "Conclusion."

MA-PCA FRAMEWORK

This section will introduce the concept of our multi-model adaptation learning with possibilistic clustering assumption for EEG-based (MA-PCA) framework in detail. It mainly uses multiple-source models which are obtained from existing relevant source datasets to learn the robust semi-supervised classification model. Therefore, the two core components are organically unified into MA-PCA: (1) any instance should have a similar label membership with its corresponding LWM. The fuzzy entropy regularization term is added to reach the amount of discrimination information to improve the classification accuracy and robustness and (2) assuming that multiple-source models can help SLL, the existing multi-source models are used for multi-source domain adaptation learning to establish a robust target domain classification model. At the same time, considering IID and non-IID, the best source model is found by multiple-source models with different weights to train the target model.

Notations

We denote $X = \{\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_i, \mathrm{x}_{i+1}, ..., \mathrm{x}_n\}$ as a feature dataset, where n is sample number ($l \ll n$) and $Y_l = \{y_1, y_2, ..., y_l\}^T \in \mathbb{R}^{l\times M}$ is a label set about dataset$X_l = \{x_i\}_{i=1}^{l}$. $X_u = \{x_j\}_{j=l+1}^{n}$ is an unlabeled feature dataset, where x_i is d dimensions ($x_i \in R^d$) vector of the i-th sample. We compute LWM $\hat{x}_i$ about x_i:

$$\hat{x}_i = \frac{\sum_{x_j \in Ks(x_i)} D_{ij} x_j}{\sum_{x_j \in Ks(x_i)} D_{ij}}, \tag{1}$$

where the k nearest neighbors of x_i are arranged in $Ks\,(x_i)$, and the Euclidean distance algorithm is used to find these neighbors. We design an undirected weight graph $G = (X, D)$, where $D \in \mathbb{R}^{n\times n}$ is weight matrix, $D_{ji} = D_{ij} \geq 0$, and the element is measured as follows:

$$D_{ij} = \begin{cases} \exp(-\tau||x_i - x_j||^2), & x_i \text{ is one of the neighbors of } x_j \\ 0 & \text{otherwise} \end{cases},$$

where τ is a changeable parameter in Gaussian kernel function. If the distance between x_i and x_j is smaller, D_{ij} is higher, and *vice versa*. Therefore, the clustering problem is changed into a graph problem in this paper.

Basic Formulation of MA-PCA

Since both SSPCA and SA-SSCCM methods require that the training data and test data meet the IID assumption, this paper reasonably combines the SSPCA method with a multi-model adaptation learning method (i.e., MA-PCA). This proposed method not only improves the robustness on noises/outliers but also solves the problems of insufficient label data and noisy data affecting the performance of the model and the different distribution of training data and test data. We therefore propose the following basic formula of MA-PCA:

$$Q\left(\boldsymbol{W}, \boldsymbol{v}_m\left(\boldsymbol{x}_j\right), \gamma\right) = \min \Omega_B(\boldsymbol{W}, \boldsymbol{v}_m\left(\boldsymbol{x}_j\right)) + \beta\Omega_M(\boldsymbol{W}, \gamma), \tag{2}$$

where $\Omega_M\left(\boldsymbol{W}, \gamma\right)$ is the multi-model adaptation term, and $\Omega_B\left(\boldsymbol{W}, v_m\left(x_j\right)\right)$ is used for reducing the negative impact of noises/outliers. We have the following function:

$$\begin{aligned} &\Omega_B\left(\boldsymbol{W}, \boldsymbol{v}_m\left(\boldsymbol{x}_j\right)\right) \\ &= \min_{w, v_m(x_j)} \sum_{i=1}^{l} ||\boldsymbol{W}^T\boldsymbol{x}_i - \boldsymbol{y}_i||^2 + \lambda_s \sum_{i=1}^{l} ||\boldsymbol{W}^T\hat{\boldsymbol{x}}_i - \boldsymbol{y}_i||^2 \\ &\quad + \sum_{m=1}^{M}\sum_{j=l+1}^{n} \boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right) ||\boldsymbol{W}^T\boldsymbol{x}_j - \boldsymbol{c}_m||^2 \\ &\quad + \lambda_s \sum_{m=1}^{M}\sum_{j=l+1}^{n} \boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right) ||\boldsymbol{W}^T\hat{\boldsymbol{x}}_j - \boldsymbol{c}_m||^2 + \lambda||\boldsymbol{W}^T||_{\mathcal{H}}^2 \\ &\quad + C\sum_{m=1}^{M}\sum_{j=l+1}^{n} \left(\boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right)\ln \boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right) - v_m^2\left(x_j\right)\right) \\ &\quad s.t. 0 \leq \boldsymbol{v}_m\left(\boldsymbol{x}_j\right) \leq 1, m = 1, \ldots, M, j = l+1, \ldots, n, \end{aligned} \tag{3}$$

where λ_s, λ, C are balance parameters that can be adjusted to avoid overfitting during model training. The details about the other parameters are provided in Dan et al. (2021).

The Multi-Model Adaptation Term in MA-PCA

In our domain adaptation learning, given is $\left\{\boldsymbol{W}_i^s\right\}_{i=1}^{q}$ as a multiple-source model set, where q is source model number, and $\boldsymbol{W}_i^s$ is the i-th source model. Each source model is obtained by learning the specified public dataset. This paper expects the

classification results of the target model to be consistent with those of the source domain models. In other words, this paper will learn the target classifier $f(x) = \boldsymbol{W}^T\boldsymbol{x}$ in the whole sample space, regularize the weight parameters of different source domain models to control its complexity, and make each target instance be close to the source domain models. According to this criterion, it can be realized by introducing the multi-source domain model adaptation regularization function on the target domain. The formula is described as follows:

$$\begin{aligned}
&\Omega_M(W,\gamma)\\
&=\sum_{i=1}^{q}\gamma_i\int_X||\boldsymbol{W}^T\boldsymbol{x}-\boldsymbol{W}_i^T\boldsymbol{x}||^2d\boldsymbol{x}+\eta||\gamma||_2^2\\
&\triangleq\sum_{i=1}^{q}\gamma_i\frac{1}{n}\sum_{j=1}^{n}||\boldsymbol{W}^T\boldsymbol{x}_j-\boldsymbol{W}_i^{sT}\boldsymbol{x}_j||_F^2+\eta||\gamma||_2^2\\
&=\sum_{i=1}^{q}\gamma_i tr\left[\left(\boldsymbol{W}-\boldsymbol{W}_i^s\right)^T\boldsymbol{S}\left(\boldsymbol{W}-\boldsymbol{W}_i^s\right)\right]+\eta||\gamma||_2^2,
\end{aligned}\tag{4}$$

where $\gamma=\left[\gamma_1,\ldots,\gamma_q\right]^T$, and γ_i is a weight of the *i-th* source model. We constraint that the sum of γ_i is 1 (i.e., $\sum_{i=1}^{q}\gamma_i=1$). It is better to explore the contribution among the source models. The divergence matrix of the target domain is $\boldsymbol{S}=\boldsymbol{X}\boldsymbol{X}^T$. $\eta\in R^+$ is a balance parameter to control the contribution of $||\gamma||_2^2$. This parameter can be changed to avoid overfitting on multiple-source models. In addition, in the second equation in (4), we employ the sampling frequency as a weight to access the real distribution for the target domain.

Remark 1: The divergence matrix $\boldsymbol{S}$ is important for connecting the source classifiers and the target classifier. It will promote the learning of the target classification model to the real distribution direction of the target domain, thereby improving the generalization performance of adaptive learning, which is essentially different from other domain adaptation regularization terms based on DAL models (Bottou and Vapnik, 1992; Duan et al., 2012a). In order to better fit the model idea, this paper refers to formula (4) as the construction of the regularization term for divergence-constrained multi-model adaptation.

Final Formulation

We expect better model adaptation performance for EEG-based emotion recognition by combining SSPCA with scatter-constrained multi-source classifier model. Therefore, a unified framework MA-PCA is obtained to learn $\boldsymbol{W},\boldsymbol{v}_m\left(\boldsymbol{x}_j\right),\gamma$ by combining formulas (3) and (4). The optimization problem of MA-PCA can be described as follows:

$$\begin{aligned}
&Q\left(\boldsymbol{W},\boldsymbol{v}_m\left(\boldsymbol{x}_j\right),\gamma\right)\\
&=\min_{w,v_m(x_j)}\sum_{i=1}^{l}||\boldsymbol{W}^T\boldsymbol{x}_i-\boldsymbol{y}_i||^2+\lambda_s\sum_{i=1}^{l}||\boldsymbol{W}^T\hat{\boldsymbol{x}}_i-\boldsymbol{y}_i||^2\\
&+\sum_{m=1}^{M}\sum_{j=l+1}^{n}\boldsymbol{v}_m^2\left(x_j\right)||\boldsymbol{W}^T\boldsymbol{x}_j-\boldsymbol{c}_m||^2\\
&+\lambda_s\sum_{m=1}^{M}\sum_{j=l+1}^{n}\boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right)||\boldsymbol{W}^T\hat{\boldsymbol{x}}_j-\boldsymbol{c}_m||^2+\lambda||\boldsymbol{W}^T||_{\mathcal{H}}^2\\
&+C\sum_{m=1}^{M}\sum_{j=l+1}^{n}\left(\boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right)\ln\boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right)-\boldsymbol{v}_m^2\left(\boldsymbol{x}_j\right)\right)\\
&+\beta\left\{\begin{array}{l}\sum_{i=1}^{q}\gamma_i tr\left[\left(\boldsymbol{W}-\boldsymbol{W}_i^s\right)^T\boldsymbol{S}\left(\boldsymbol{W}-\boldsymbol{W}_i^s\right)\right]\\+\eta||\gamma||_2^2\end{array}\right\},
\end{aligned}\tag{5}$$

where $\gamma=[\gamma_1,\gamma_2,\ldots\gamma_q]^T$, $\gamma\in\mathbb{R}^{q\times1}$, $\gamma^T\boldsymbol{1}_q=1$, and q are the number of source domain models. When $\beta=0$, MA-PCA degenerated to SSPCA. When $\beta>0$, β is used as a balance parameter. When γ_i is constant 1, it indicates that there is only one single-source domain, and its distribution is the same as the target domain. At this time, MA-PCA approximates the SA-SSCCM method. When $0\le\gamma_i<1$, MA-PCA is a multi-model adaptation learning method based on the possibility clustering assumption, and the distribution of source domain and target domain can be identical or non-identical.

OPTIMIZATION

The objective function (5) is a non-convex function on $(\boldsymbol{W},\boldsymbol{v}_m\left(\boldsymbol{x}_j\right),\gamma)$. In this paper, the strategy of alternating iterative optimization is adopted to realize the optimal solution of decision model $\boldsymbol{W}$, membership model $\boldsymbol{v}_m\left(\boldsymbol{x}_j\right)$, and contribution coefficient γ of the source models, respectively, and each iteration has an optimal solution.

Update $\boldsymbol{W}$ as Given $\boldsymbol{v}_m\left(\boldsymbol{x}_j\right)$ and γ

Firstly, fixing $\boldsymbol{v}_m\left(\boldsymbol{x}_j\right)$ and γ to solve $\boldsymbol{W}$: for ease of calculation, the following formula (5) is transformed into matrix form. $\boldsymbol{W}$ can be written as $\boldsymbol{W}=\sum_{i=1}^{n}x_i\alpha_i$. Under the Representation Theorem, the formula (5) exists in reproducing kernel Hilbert space, and the kernel of W can be rewritten as follows: $W=\sum_{i=1}^{n}K(x_i,x)\alpha_i$ (Belkin et al., 2006). Therefore, formula (5) is mapped into a finite dimensional space of the optimization α_i and can be reformulated as follows:

$$\begin{aligned}
Q(\alpha)=\min_{\alpha}\;&tr\left(\left(\boldsymbol{\alpha}^T\boldsymbol{K}_l\boldsymbol{K}_l-Y\right)\left(\boldsymbol{\alpha}^T\boldsymbol{K}_l\boldsymbol{K}_l-Y\right)^T\right)\\
&+\lambda_s tr\left(\left(\boldsymbol{\alpha}^T\overline{\boldsymbol{K}_l\boldsymbol{K}_l}-\boldsymbol{Y}\right)\left(\alpha^T\overline{\boldsymbol{K}_l\boldsymbol{K}_l}-Y\right)^T\right)\\
&+tr\left(\left(\boldsymbol{\alpha}^T\boldsymbol{K}_u\boldsymbol{K}_u\boldsymbol{J}-\boldsymbol{L}\right)\hat{\boldsymbol{V}}\left(\boldsymbol{\alpha}^T\boldsymbol{K}_u\boldsymbol{K}_u\boldsymbol{J}-\boldsymbol{L}\right)^T\right)\\
&+\lambda_s tr\left(\left(\boldsymbol{\alpha}^T\overline{\boldsymbol{K}_u\boldsymbol{K}_u}\boldsymbol{J}-\boldsymbol{L}\right)\hat{\boldsymbol{V}}\left(\boldsymbol{\alpha}^T\overline{\boldsymbol{K}_u\boldsymbol{K}_u}\boldsymbol{J}-\boldsymbol{L}\right)^T\right)\\
&+\lambda tr\left(\boldsymbol{\alpha}^T\boldsymbol{K}\boldsymbol{K}\boldsymbol{\alpha}\right)+\beta\sum_{i=1}^{a}\gamma_i\left(\boldsymbol{W}_I^T-\boldsymbol{\alpha}^T\boldsymbol{K}\right)\boldsymbol{K}\\
&\cdot\boldsymbol{K}\left(\boldsymbol{W}_i-\boldsymbol{K}\boldsymbol{\alpha}\right),
\end{aligned}\tag{6}$$

where the details about $\boldsymbol{J}$, $\boldsymbol{L}$, $\hat{\boldsymbol{V}}$ can be found in Dan et al. (2021).

By solving the derivation of (6) w.r.t. α and letting it be equal to 0 (i.e., $\partial Q(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha} = 0$), we obtain the following:

$$\boldsymbol{\alpha} = \boldsymbol{H}^{-1}\boldsymbol{B}, \tag{7}$$

where

$$\begin{aligned}\boldsymbol{H} &= \boldsymbol{K}_l\boldsymbol{K}_l\boldsymbol{K}_l\boldsymbol{K}_l + \boldsymbol{\lambda}\overline{\boldsymbol{K}_l\boldsymbol{K}_l\boldsymbol{K}_l\boldsymbol{K}_l} \\ &+ \boldsymbol{K}_u\boldsymbol{K}_u\boldsymbol{J}\hat{\boldsymbol{V}}\boldsymbol{J}^T\boldsymbol{K}_u\boldsymbol{K}_u \\ &+ \boldsymbol{\lambda}_s\overline{\boldsymbol{K}_u\boldsymbol{K}_u}\boldsymbol{J}\hat{\boldsymbol{V}}\boldsymbol{J}^T\overline{\boldsymbol{K}_u\boldsymbol{K}_u} + \boldsymbol{\lambda}\boldsymbol{K}\boldsymbol{K} + \boldsymbol{\beta}\sum_{i=1}^{q}\gamma_i\boldsymbol{K}\boldsymbol{K}\boldsymbol{K}\boldsymbol{K},\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{B} &= \boldsymbol{K}_l\boldsymbol{K}_l\boldsymbol{Y}^T + \lambda_s\overline{\boldsymbol{K}_l\boldsymbol{K}_l}\boldsymbol{Y}^T + \boldsymbol{K}_u\boldsymbol{K}_u\boldsymbol{J}\hat{\boldsymbol{V}}\boldsymbol{L} \\ &+ \boldsymbol{\lambda}_s\overline{\boldsymbol{K}_u\boldsymbol{K}_u}\boldsymbol{J}\hat{\boldsymbol{V}}\boldsymbol{L}^T + \boldsymbol{\beta}\sum_{i=1}^{q}\boldsymbol{\gamma}_i,\ \boldsymbol{K}\boldsymbol{K}\boldsymbol{K}\boldsymbol{W}_i^s,\end{aligned}$$

Finally, the solution of $\boldsymbol{W}$ is $\boldsymbol{K}\boldsymbol{\alpha}$.

Update $\boldsymbol{v_m}(\boldsymbol{x_j})$ as Given $\boldsymbol{W}$ and $\boldsymbol{\gamma}$

Then, we fix $\boldsymbol{W}$ and γ to solve $\boldsymbol{v}_m(\boldsymbol{x}_j)$. The optimal problem of the objective function in (5) can be rewritten as follows:

$$\begin{aligned}&Q\left(\boldsymbol{v}_m(\boldsymbol{x}_j)\right) \\ &= \min_{v_m(x_j)}\sum_{m=1}^{M}\sum_{j=l+1}^{n} v_m\left(x_j\right)^2||\boldsymbol{W}^T\boldsymbol{x}_j - \boldsymbol{c}_m||^2 \\ &+ \lambda_s\sum_{m=1}^{M}\sum_{j=l+1}^{n} v_k\left(x_j\right)^2||\boldsymbol{W}^T\hat{\boldsymbol{x}}_j - \boldsymbol{c}_m||^2 + \lambda||\boldsymbol{W}^T||_{\mathcal{H}}^2 \\ &+ C\sum_{m=1}^{M}\sum_{j=l+1}^{n}(\boldsymbol{v}_m(\boldsymbol{x}_j)^2\ln\boldsymbol{v}_m(\boldsymbol{x}_j)^2 - \boldsymbol{v}_m(\boldsymbol{x}_j)^2),\end{aligned} \tag{8}$$

By solving the derivation of Eq. (5) w.r.t. $\boldsymbol{v}_m(\boldsymbol{x}_j)$ and letting it be equal to zero (i.e., $\partial Q\left(\boldsymbol{v}_m(\boldsymbol{x}_j)\right)/\partial v_m(x_j) = 0$), then

$$\begin{aligned}\partial Q\left(\boldsymbol{v}_m(\boldsymbol{x}_j)\right)/\partial \boldsymbol{v}_m(\boldsymbol{x}_j) &= 2\boldsymbol{v}_m(\boldsymbol{x}_j)||\boldsymbol{W}^T\boldsymbol{x}_j - c_m||^2 \\ &+ 2\lambda_s\boldsymbol{v}_m(\boldsymbol{x}_j)||\boldsymbol{W}^T\hat{x}_j - c_m||^2 \\ &+ C\left[2\boldsymbol{v}_m(\boldsymbol{x}_j)\log\boldsymbol{v}_m\left(\boldsymbol{x}_j\right)^2\right] = 0,\end{aligned}$$

we can get:

$$\boldsymbol{v}_m(\boldsymbol{x}_j) = \exp\left(\frac{-(||\boldsymbol{W}^T\boldsymbol{x}_j - c_m||^2 + ||\boldsymbol{W}^T\hat{x}_j - c_m||^2)}{2C}\right), \tag{9}$$

Since $\boldsymbol{x}_j$ means any one instance, the general presentation of $\boldsymbol{v}_m(\boldsymbol{x}_j)$ is as follows:

$$\boldsymbol{v}_m(\boldsymbol{x}) = \exp\left(\frac{-\left(||\boldsymbol{W}^T\boldsymbol{x} - \boldsymbol{c}_m||^2 + ||\boldsymbol{W}^T\hat{\boldsymbol{x}} - \boldsymbol{c}_m||^2\right)}{2C}\right), \tag{10}$$

Update $\boldsymbol{\gamma}$ by Fixing $\boldsymbol{W}$ and $\boldsymbol{v_m}(\boldsymbol{x_j})$

We define $\boldsymbol{A}_i = tr\left[\left(\boldsymbol{W} - \boldsymbol{W}_i^s\right)^T\boldsymbol{S}\left(\boldsymbol{W} - \boldsymbol{W}_i^s\right)\right](i = 1, ..., q)$ which is corresponding to the adaptation of the i-th source model to the target domain. Given $\boldsymbol{W}$ and $\boldsymbol{v}_m(\boldsymbol{x}_j)$, the objective function (5) can be rewritten in (11) as follows:

$$\begin{aligned}&\min_{\gamma}\gamma^T\boldsymbol{A} + \eta||\gamma||_2^2 \\ &s.t.\gamma^T\boldsymbol{1}_q = 1,\ 0 \le \gamma \le 1,\end{aligned} \tag{11}$$

where $\boldsymbol{A} = \left(\boldsymbol{A}_1, ..., \boldsymbol{A}_q\right)^T$. The optimal estimation of γ becomes the optimal weight division problem of multiple model adaptation learning with scatter constraints. Theoretically, if $\eta = 0$, the optimal γ_i will be 1; otherwise, if $\boldsymbol{A}_i = \min_{j=1,...,q}\boldsymbol{A}_j$, γ_i is 0; if $\eta \to +\infty$, the optimal γ_i will tend to be the same weight $1/q$. Thus, the following theorem is obtained:

Theorem 1 (Karasuyama and Mamitsuka, 2013). The following equation is the optimal solution of (11):

$$\gamma_i = \begin{cases}\frac{\rho - \boldsymbol{A}_i}{\eta}, & i = 1, 2, ..., \zeta \\ 0, & i = \zeta + 1, ..., q\end{cases},$$

where $\rho = \left(\eta + \sum_{i=1}^{q}\boldsymbol{A}_i\right)/\zeta$, $\zeta = \left|\left\{i|\rho - \boldsymbol{A}_i > 0, i = 1, 2, ..., q\right\}\right|$.

Theorem 1 presents that there are ζnon-zero entries in the optimal γ. According to the target domain, this optimization can select source domains automatically with a different γ. If γis bigger, there is a higher similarity between the source domain and the target domain. Since the optimal γcould be calculated *via* the optimal ζ, Karasuyama and Mamitsuka, 2013) presented a special method to obtain the optimal ζ. This algorithm effectiveness can be proved under the given right amount of source domains.

Next, the algorithm adopts a coordinate descent strategy to solve (11), which is close to the method in Geng et al. (2012). In each iteration cycle, when other entries are fixed, we select two items to update, and $\gamma^T 1_q = 1$ must be satisfied at the end of each iteration. Suppose that in an iteration cycle the i-th and j-th entries are selected, the following iterative formula can therefore be obtained:

$$\begin{cases}\text{(I) } \gamma_i^* = 0, \gamma_j^* = \gamma_i + \gamma_j, \textit{ if } 2\eta\left(\gamma_i + \gamma_j\right) + \left(\boldsymbol{A}_j - \boldsymbol{A}_i\right) \le 0 \\ \text{(II) } \gamma_j^* = 0, \gamma_i^* = \gamma_i + \gamma_j, \textit{ if } 2\eta\left(\gamma_i + \gamma_j\right) + \left(\boldsymbol{A}_i - \boldsymbol{A}_j\right) \le 0 \\ \text{(III) } \textit{or } \gamma_i^* = 1/2\left(\gamma_i + \gamma_j\right) + 1/4\eta\left(\boldsymbol{A}_j - \boldsymbol{A}_i\right), \\ \quad \gamma_j^* = \gamma_i + \gamma_j - \gamma_i^*.\end{cases}, \tag{12}$$

We iteratively traverse all paired entries in γ and optimize any two entries in γ by (12) until the optimization function (5) converges. Intuitively, the updating criteria in (12) tends to be that the larger the value to γ_i, the smaller the $\boldsymbol{A}_i$. Since $\boldsymbol{A}_i$ measures the distribution distance between the i-th source model and the target domain, the smaller the $\boldsymbol{A}_i$, the higher the correlation between the i-th source domain and the target domain.

Remark 1: After obtaining the optimal solution of W and V, the label matrix of the samples in the target domain can also be obtained, and the influence of noise has been effectively suppressed. The label and the membership of out of sample can also be calculated by W and V. Finally, the performance of W relies on the learned graph $G = XD$ and the multiple-source domain adaptation model.

Convergence Analysis

Since the objective function (5) is a multi-objective optimization function, it is difficult to guarantee the global optimal solution. It is worthy to note that the algorithm in this paper adopts an alternate iterative strategy for optimizing. Since the objective function of a single optimization variable [i.e., equations (6), (8), and (11)] is convex and in closed form, the iterative optimal solution can be obtained. Therefore, this work only deduces the asymptotic convergence of the algorithm based on the iterative target value of the objective function. The derivation process is shown in (13) as follows:

$$\begin{aligned} &Q\left(W^{\mathrm{itr}}, v^{itr}, \gamma^{itr}\right) \geq \\ &Q\left(W^{itr+1}, v^{itr}, \gamma^{itr}\right) \geq \\ &Q\left(W^{itr+1}, v^{itr+1}, \gamma^{itr}\right) \geq \\ &Q\left(W^{itr+1}, v^{itr+1}, \gamma^{itr+1}\right) > \varepsilon > 0, \end{aligned} \tag{13}$$

where $\mathrm{W}^{\mathrm{itr}}, \mathrm{v}^{itr}, \gamma^{itr}$ are the optimal solutions at the *itr*-th iteration. ε is a very small constant. The objective function will converge to a local optimum. The derivation process proves that the iterative target value of the algorithm shows a downward trend. When the value drops to a certain threshold (at least greater than ε), we stop the iterative. Finally, the objective function will converge to the local target value of each single optimization variable.

ALGORITHM DESCRIPTION

The optimization of MA-PCA adopts an alternating iteration strategy. The most semi-supervised learning methods are often optimized by iterative learning. In addition, the membership value of initialized unlabeled instances can be obtained by any of the following methods: some fuzzy clustering method, randomization strategy, or all initialized to zero. Therefore, the learning of MA-PCA method starts with labeled instances to initialize the decision model W. When the objective function converges, the iteration terminates. The algorithm in this paper specifically adopts a window-based stopping criterion to better control the algorithm convergence: given a window size $\hbar$, computing $\varsigma = |Max\Theta_{itr} - Min\Theta_{itr}|/Max\Theta_{itr}$ at the *itr*-th iteration ($\Theta_{itr} = \{Obj_{itr-\hbar+1}, ..., Obj_{itr}\}$means $\Theta_{itr}$is composited by the historical target value in this window. When $\varsigma < \varepsilon$, the iteration terminates. The details of this algorithm are shown in **Table 1**.

TABLE 1 | Algorithm description of MA-PCA.

Input: the target domain with data $\boldsymbol{X}_l$ and its labels $\boldsymbol{Y}_l$, unlabeled data $\boldsymbol{X}_u$, regular term parameter λ, λ_s, C, η, β. There are q source classifier models $\{\boldsymbol{W}_i^s\}_{i=1}^q$, iteration termination threshold ε, and maximum number of iterations N.

Output: the target classifier model $\boldsymbol{W}$, the label membership function $\boldsymbol{v}$, and the contribution coefficient $\boldsymbol{\gamma}$ of multi-source model.

Procedure:

1. Initialize the label memberships of unlabeled data, $\gamma_i^0 = 1/\mathrm{q}\,(\mathrm{i} = 1, ..., q)$;
2. Obtain the initial $\boldsymbol{W}^0$ by Eq. (7);
3. Obtain the initial $\boldsymbol{v}^0$ by Eq. (10);
4. Calculate the $Q\left(\boldsymbol{W}^0, v^0, \gamma^0\right)$ of objective function

for *itr* = 1 **to** *N* **do**

{

5.1 Fix the current $\boldsymbol{v}^{itr}$ and γ^{itr} for updating $\boldsymbol{W}^{itr}$ to $\boldsymbol{W}^{itr+1}$ by Eq. (7)

5.2 Fix the current $\boldsymbol{W}^{itr}$ and γ^{itr} for updating $\boldsymbol{v}^{itr}$ to $\boldsymbol{v}^{itr+1}$ by Eq. (10)

5.3 Fix the current $\boldsymbol{W}^{itr}$ and $\boldsymbol{v}^{itr}$ for updating γ^{itr} to γ^{itr+1} by Eq. (12)

Until $itr > N$ or $\varsigma < \varepsilon$, return the optimal $\boldsymbol{W}$, $\boldsymbol{v}$, and γ

}

EXPERIMENT

In this part, we comprehensively compare the proposed method with several state-of-the-arts on two widely used benchmark databases, including SEED (Zheng and Lu, 2015) and DEAP (Koelstra et al., 2012), for EEG-based emotion recognition (Mansour et al., 2009).

Datasets

According to Lan et al. (2018) and Zhong et al. (2020), there exist certain significant differences between SEED and DEAP since they can be generated by different subjects, sessions, EEG devices, experimental schemes, emotional stimuli, etc., Detailed information about these two datasets can be viewed in Lan et al. (2018). In the following experiments, we adopt differential entropy (Lan et al., 2018; Zhong et al., 2020) as the data feature in emotion recognition, which has also been widely used in the preceding literatures (Shi et al., 2015; Zheng et al., 2015, 2016; Chai et al., 2016, 2017; Lan et al., 2018; Zhong et al., 2020) for domain adaptation emotion recognition.

Baseline Setting

We will systematically compare our method with such state-of-the-arts as SSPCA (Dan et al., 2021), a baseline without domain adaptation, FastDAM (Duan et al., 2012b), Multi-KT (Tommasi et al., 2014) with l_2-norm constraint on p, A-SVM (Yang et al., 2007), and DSM (Duan et al., 2012a). Since existing deep domain adaptation frameworks have achieved many inspiring results on emotion recognition as well as visual recognition, we also additionally present comparisons with several deep (CNN-based) domain adaptation methods with deep features: DAN (Long et al., 2015) and Reverse Grad (Ganin and Lempitsky, 2015).

It was noted that, in the DA schema, automatic parameter tuning is not possible for source classifiers using cross-validation due to the reason that training and test data are from different

data distributions. Therefore, all methods evaluated in this paper on the dataset are empirically searched for the parameter space to optimize parameter settings to obtain the best results for each method. The parameters of all methods are adjusted to obtain the best results, except for the specially specified parameters.

For the method SSPCA without domain adaptation, this experiment will fuse the decision values of all classifiers obtained from the independent training of labeled samples in each source domain and target domain. DSM, Multi-KT, A-SVM, and FastDAM are domain adaptation methods. For A-SVM, this experiment also fuses the decision values of all basic classifiers, and each classifier is learned from a labeled sample in a source domain.[1]

There are some hyper-parameters in the objective function (5) that need to be determined. First of all, this experiment sets the most important parameters (e.g., γ_i) as optimization variables in the iterative optimization process, and only a few crucial parameters that are proposed in MA-PCA need to be pre-defined (e.g., λ, λ_s, C, η, β), and considering that parameter determination is an open problem in the field of machine learning, we have determined parameters empirically in a past work. Since the exponent of γ_i plays the role of avoiding trivial solutions in the process of optimizing γ_i, as proved in Hou et al. (2017), the larger the exponent of γ_i is, the closer all weight values are to be same. In order to reflect the differences among different source domains, this experiment will set the index of γ_i to be 2 by experience. The validity of this decision will be verified in the experimental results in the following section. The hyperparameter λ, λ_s, C, η, β is adjusted within the range of $\{10^{-4}, 10^{-3}, ..., 10^{3}, 10^{4}\}$. Finally, we search the nearest neighbor number k from the set $\{35, 10, 15, 17\}$ to construct the nearest neighbor graph in MA-PCA (also SSPCA) and obtain the first-ranked recognition accuracy from the optimal parameter configuration.

For non-linear learning methods MA-PCA, FastDAM, and multi-KT, Gaussian kernel $K_{i,j} = \exp\left(-\sigma||x_i - \boldsymbol{x}_j||^2\right)$ is default kernel function, $\sigma = {}^{\mathbf{1}}/_{d}$, and d is the feature dimension. In FastDAM, $\gamma_i = \frac{\exp(-\delta Dist(X_i^s, X))}{\sum_i \exp(-\delta Dist(X_i^s, X))}$ $(i = 1, ..., S)$ is the weight value of each source domain, $\delta = 100$. For the benchmark method SSPCA, the target domain samples are directly mapped to the source domain without any domain adaptation, and the decision values of all classifiers obtained from the independent training of labeled samples from each source domain and target domain are equally fused.

Emotion Recognition Within Dataset

Note that different subjects even from the same dataset still have different EEG feature distributions due to individual characteristics. We therefore practice the so-called leave-one-out cross-validation strategy conducted also in Lan et al. (2018) to evaluate the emotion recognition performance of MA-PCA—that is, one subject remained to be the target domain, and the others from the dataset are constructed as multiple sources. In this multi-source scenario, we follow the same setting as Tao et al. (2021) to evaluate our method compared with other state-of-the-arts on SEED and DEAP, respectively.

Performance Comparison

The emotion recognition performance of MA-PCA and the rest of the comparison methods within the DEAP and SEED datasets are visualized in **Figure 1**. It can be seen from the bar plot that the recognition performance of all DA methods is better than that of SSPCA, and MA-PCA achieves the best performance (about 21% performance improvement over SSPCA), followed by DSM on DEAP dataset. Besides this, those multi-source adaptation methods, including our method, unsurprisingly achieved more accuracy gains than the no-adaptation method SSPCA on SEED. We can observe that our method MA-PCA demonstrates the best performance on SEED by upgrading the average accuracy. An interesting observation is that all methods work better on SEED than on DEAP, which has also been reported in Lan et al. (2018) and Tao et al. (2021). The reason for this phenomenon might be that the larger distribution discrepancy between different subjects from DEAP prevented boosting performance in these methods (Mansour et al., 2009; Lan et al., 2018).

Finally, MA-PCA achieved almost the best performance on both datasets. A possible explanation is that the distribution discrepancy may exist in the same dataset (i.e., DEAP or SEED), and MA-PCA can learn a more robust target classifier for domain adaptation by discriminatively selecting a set of prelearned base classifiers in the non-IID scenario of multi-subject adaptation.

Multi-Kernel Learning

As well known, the choice of kernel is a challenging issue in the kernel learning method. Recently, multiple kernel learning (MKL) has been effectively proposed for conquering this choice issue that existed in single kernel learning methods. Consequently, we also evaluate the performance boost in our method by using MKL (called MKMA-PCA for short) for each source domain. To this end, the first step is to construct a new space spanned by multiple kernel mapping features. We firstly denote by $\{\phi_a\}_{a=1}^{\mho}$ an empirical kernel function set, which, respectively, projected X_a into $\mho$ different spaces. Then, an orthogonally integrated space can be constructed by concatenating these $\mho$ spaces. We denote the mapping features in this final space by $\widetilde{\phi}(x_i) = [\phi_1(x_i)^T, \phi_2(x_i)^T, ..., \phi_{\mho}(x_i)^T]^T \in \mathbb{R}^{\mho n_a}$, where $x_i \in X_a$. Correspondingly, the kernel matrix in this final space can be easily deduced as $K_{new} = [\widetilde{K}_1; \widetilde{K}_2; ...; \widetilde{K}_{\mho}]$, where $\widetilde{K}_i$ is the i-th kernel matrix from the $\mho$ feature spaces. Aiming to exploit the multiple kernel spaces, we therefore employ four kernel mapping functions, including the Gaussian kernel used above. The other additionally employed kernels are inverse square distance kernel function, Laplacian kernel function, and inverse distance kernel function, respectively, denoted as $\mathbf{K}_{ij} = 1/\left(1 + \sigma||x_i - \boldsymbol{x}_j||^2\right)$, $\mathbf{K}_{ij} = \exp\left(-\sqrt{\sigma}||x_i - \boldsymbol{x}_j||\right)$, and $\mathbf{K}_{ij} = 1/(1 + \sqrt{\sigma}||x_i - \boldsymbol{x}_j||)$.

[1]For each source domain, we train one SVM by using the corresponding labeled samples. Then, for each test instance x, the decision values from p SVM classifiers are converted into probability values by using the sigmoid function [i.e., g(t) = 1/(1 + exp(-t)]. Finally, we average the p probability values as the final prediction of the test instance x.

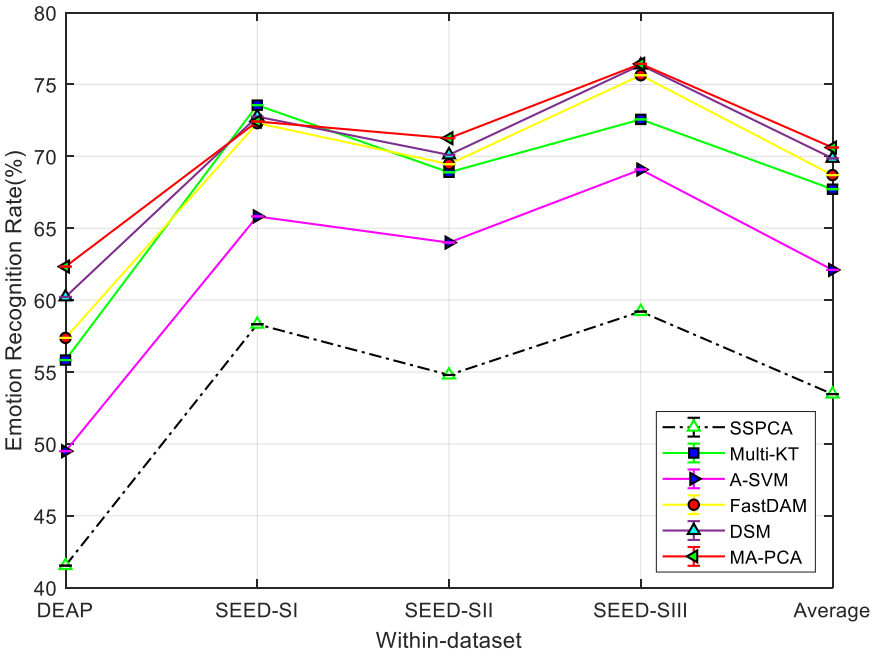


FIGURE 1 | Domain adaptation emotion recognition on within dataset. SI, session I; SII, session II; SIII, session III.

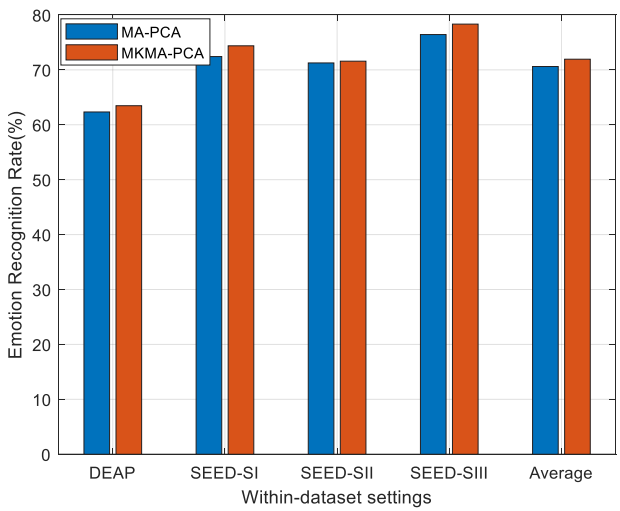


FIGURE 2 | Emotion recognition on within-dataset with multiple kernel learning. SI, session I; SII, session II; SIII, session III (similarly hereinafter).

The observation from **Figure 2**, in which MKMA-PCA significantly outperforms MA-PCA, justifies that our MA-PCA with MKL can further boost the recognition performance on DEAP and SEED. This also proves the importance of kernel choice in those kernel-based learning models.

Emotion Recognition Cross-Dataset

It is more challenging on emotion recognition when across datasets, with the differences in acquisition pathways, subjects' characteristics, and behaviors. The previous experiments show the performance comparison of MA-PCA with other DA methods within the dataset (i.e., across subjects). This subsection further evaluates the robust effectiveness of MA-PCA when adapting across datasets. In this experimental scenario, multiple different protocols were constructed using different EEG devices and emotional stimuli by sampling the training and testing datasets separately. Therefore, six experimental settings, namely, DEAP → session I, DEAP → session II, DEAP → session III, session I → DEAP, session II → DEAP, and session III → DEAP, were set up to demonstrate that MA-PCA has robust effectiveness on emotion recognition with cross-dataset. For simplicity of expression, session I, session II, and session III in SEED are coded as SI, SII, and SIII, respectively, [for detailed experimental setup information, see the literature Tao et al. (2021)].

Performance Comparison

We aim to evaluate the performance of our method MA-PCA using the emotion recognition results on cross-dataset from DEAP and SEED. The experimental results are, respectively, plotted in **Figure 3**, which shows the average results for six possible combinations.

It can be seen from the bars in **Figure 3** that the no-adaptation method SSPCA has the worst performance than others in all cases, which witnesses the existence of distribution discrepancy between DEAP and one of the sessions from SEED. In this context, the importance of domain adaptation (DA) will be indispensable. This is justified by the observation in **Figure 3** that DA may reduce technical differences in cross-dataset applications, and our MA-PCA consistently outperformed other DA methods in most cases of the cross-dataset settings. A possible

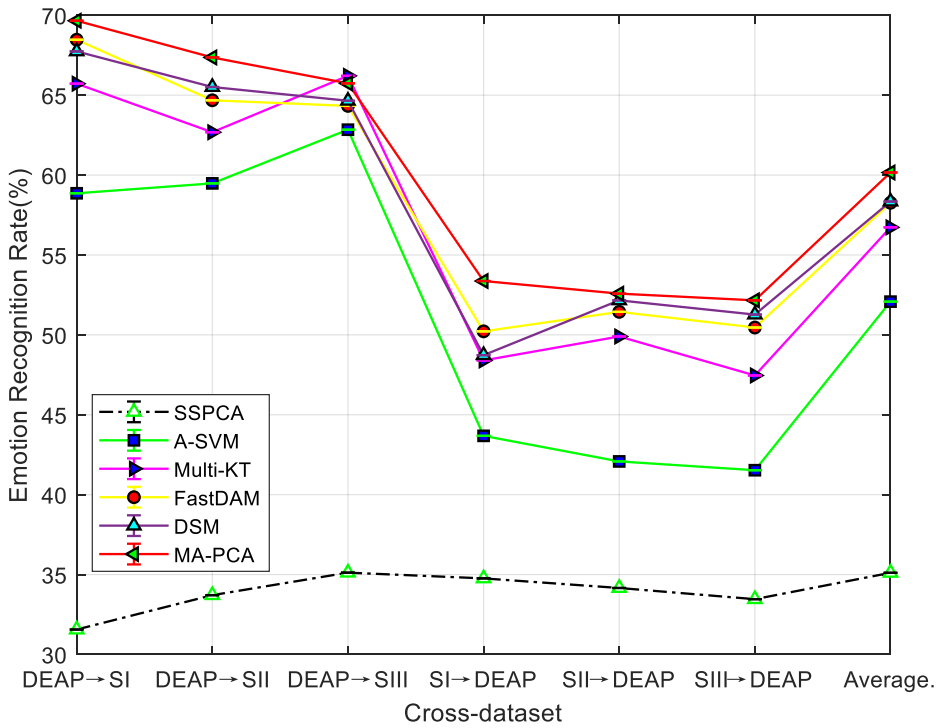


FIGURE 3 | Domain adaptation emotion recognition on cross-dataset.

reason for this may be that the clustering hypothesis with fuzzy entropy in our MA-PCA could weaken the impact of noise from these datasets. A clear phenomenon can be observed in **Figure 3**—due to the large distribution difference between the different datasets, the average recognition accuracy of all methods is correspondingly lower than the results obtained in **Figure 1** within the dataset.

Emotion Recognition With Multi-Source Prior

As reported in preceding works about domain adaptation learning, multiple-source domains can improve the adaptation performance to some extent by integrating multiple prior knowledge. Nevertheless, in concrete applications, multi-source adaptation also incurs another challenge, i.e., source scalability issue, since multi-source learning could lead to the so-called negative transfer problem. In this scenario, how to discriminately exploit multiple sources becomes a challenge worthy to be addressed in multi-source adaptation learning frameworks. To this end, we will explore in this part the different reliabilities of the prior sources in the emotion recognition task (Tao et al., 2019). We evaluate the performance of all baseline domain adaptation methods with multiple prior sources on the designed cross-dataset settings. The average accuracies of all methods are plotted in **Figure 4**, where A-SVM employs the average prior model.

When there exists a very large distribution discrepancy between different domain datasets, it is hard for A-SVM to eliminate the inter-domain distribution bias. Therefore, the results in **Figure 4** shows that A-SVM is inferior compared with the other multi-source adaptation methods in most settings. A-SVM even has a downgraded performance tendency with the increase of source domains in some scenarios, which indicates the existence of "negative transfer" phenomenon in A-SVM. Another interesting observation from **Figure 4** is that all DA methods, except A-SVM, achieve more improvement by leveraging multiple-source knowledge than that by bridging only one source (i.e., cross-subject settings) when the number of source domains increases. This proves that it is beneficial to leverage multiple sources for boosting the recognition performance. Moreover, MA-PCA and DSM conquer others by touching on the top performance due to their designed weights for discriminately screening the optimal sources. Our method MA-PCA obtains more gains over DSM in some scenarios. This may be attributed to the adopted strategy in MA-PCA, which can efficiently select the most relevant source domains through optimal weighted multi-source adaptive regularization.

Adaptive Emotion Recognition With Deeply Extracted Features

In this subsection, we will particularly evaluate our method MA-PCA with deeply extracted features by comparing it with several recently proposed deep adaptation models on cross-dataset emotion recognition using multi-source settings.

In practical tasks, our method MA-PCA can be trained on the deeply transformed features of all domains, which follows the same setup with that in Zhu et al. (2017) and Zhou et al. (2018). Concretely, some pre-trained deep models (e.g., VGG16 and DAN, etc.) are first fine-tuned using the source domain, then the deep features can be extracted from EEG signals in both

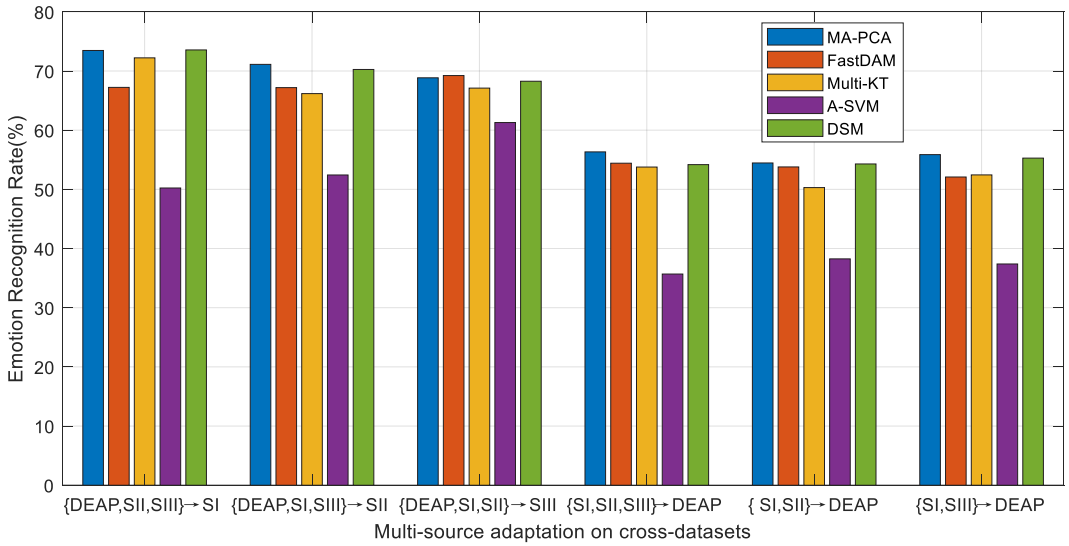


FIGURE 4 | Emotion recognition with multi-source adaptation settings.

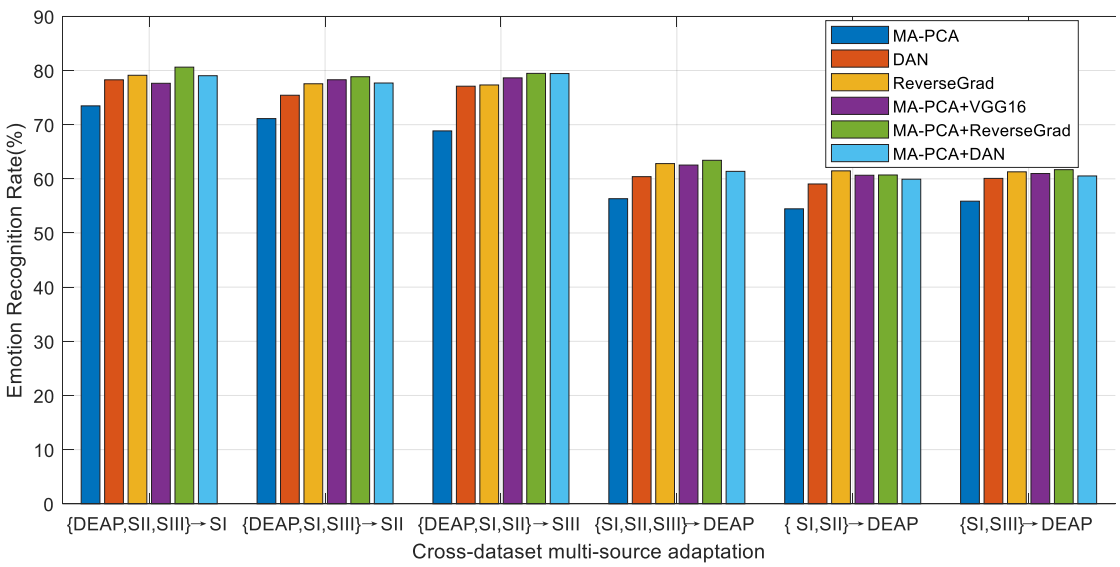


FIGURE 5 | Adaptive emotion recognition using deeply extracted features.

TABLE 2 | Multi-source adaptation emotion recognition accuracies of derived methods as well as MA-PCA.

Method	{DEAP,SII,SIII} →SI	{DEAP,SI,SIII} →SII	{DEAP,SI,SII} →SIII	{SI,SII,SIII} →DEAP	{SI,SII} →DEAP	{SI,SIII} →DEAP
MA-PCA_NTS	72.81	70.52	68.57	55.90	54.20	55.81
MA-PCA_NSS	71.30	70.05	65.87	53.17	53.77	55.43
MA-PCA-NOS	71.61	69.86	66.28	53.49	54.23	55.66
MA-PCA	**73.47**	**71.12**	**68.85**	**56.33**	**54.46**	**55.87**

Values in bold denote the best recognition rates.
SI, session I; SII, session II; SIII, session III.

source and target domains with this CNN model, and finally the recognition model would be trained on these extracted features. In the context of our experiments, we denote our methods with VGG16 (respectively DAN) model as MA-PCA + VGG16 (MA-PCA + DAN respectively). As for DAN and ReverseGrad, we use their released source codes to fine-tune the pre-trained models by, respectively, using the pre-tuned parameters in their works (Ganin and Lempitsky, 2015; Long et al., 2015). Note that these deep adaptation methods typically aim to learn domain-invariant representations. Differently from the deep adaptation frameworks, our proposed method explores to learn a domain-invariant recognition model with strong generalization ability from the source domain to the target one.

We plot the mean results of all methods in **Figure 5**, from which we can observe that our deep adaptation method MA-PCA + VGG16 and MA-PCA + DAN significantly outperform MA-PCA. This indicates the advantage of deep features due to its robust feature representation. Furthermore, MA-PCA + DAN also obtains comparable recognition performance with respect to other deep adaptation methods. This may be attributed to the classification-level constraint in MA-PCA, where most of the source discriminative structures are expected to be preserved by the guidance of target classification. In some cases, shown in **Figure 5**, MA-PCA + VGG16 or MA-PCA + DAN even achieves the top-one performance compared with other deep adaptation frameworks. This phenomenon shows that the proposed MA-PCA can become an effective surrogate to the deep adaptation model by just exploiting the deep features extracted from any one of the state-of-the-art deep models.

Ablation Study

In our method MA-PCA, there exist several hyper-parameters needed to be tuned. These hyper-parameters are mainly used to trade off different components of the proposed framework. We therefore, respectively, set these parameters into their extreme values to explore the importance of each component in MA-PCA. To this end, we set $S = \mathrm{I}$ to denote MA-PCA without target domain divergence information by MA-PCA_NTS and set $\gamma_i = 1/q$ and $\gamma_i = \frac{\exp(-Dist(X_i^s, X))}{\sum_i \exp(-Dist(X_i^s, X))} (i = 1, 2, ..., q)$to, respectively, denote by MA-PCA_NSS and MA-PCA_NOS the case where MA-PCA weights each source model by mean components and measures its distance from the target domain, respectively.

The performance of these derived methods is evaluated on cross-dataset recognition tasks, and the performance results are shown in **Table 2**. It is easy to see from **Table 2** that the performance of all derived methods is more or less degraded, and the performance of the MA-PCA_NTS method without target divergence constraints is slightly degraded. The performance of the MA-PCA_NSS method with an average weight on the source domain model decreases significantly; The performance of MA-PCA_NOS with the distance-weighted method for the source domain model is better than that of the MA-PCA_NSS method. However, the performance of these two derived methods is weaker than that of the MA-PCA optimization weighting method for the source domain model, which indicates that the proposed optimization mechanism for source domain model selection is effective. In addition, an interesting observation is that the overall recognition accuracy is below 60% when the multi-source domains are all from the SEED dataset. However, when the multi-source domain has the DEAP dataset, the overall recognition accuracy of all methods is close to 70%, even higher than 70%. It indicates that the diversity of source domains can improve the robustness and generalization of MA-PCA and its derived methods.

CONCLUSION

To deal with cross-subject/dataset EEG-based emotion recognition task, we proposed a multi-model adaptation method with possibilistic clustering assumption, i.e., MA-PCA, by exploiting the knowledge of the correlation between the source and target domains in the objective function. It suppresses the influence of noise/abnormal data and weakens the impact of model performance caused by the different distributions of training data and test data (i.e., source domain and target domain). In MA-PCA, the fuzzy entropy regularization term is used to weaken the influence of noisy data, and multi-domain adaptation learning method is used to establish a robust classification model to weaken the influence of different data distributions. The comprehensive experiments performed on two public datasets verify the effectiveness of MA-PCA in dealing with cross-subject/dataset emotion recognition. In most scenarios, our MA-PCA (or MA-PCA-VGG16/DAN) obtains the best results or comparable performance with respect to several representative baselines. Since the implementation of MA-PCA algorithm needs an iterative optimization procedure, how to improve the efficiency of MA-PCA and seek a more efficient algorithm would be an issue worthy of further study in our future 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 author.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by the research institution special program of Ningbo Polytechnic under Grant No. NZ21JG006 and Open Project of Key Laboratory of 3D Printing Equipment and Manufacturing in Colleges and Universities of Fujian Province under Grant No. 3DPM1901.

REFERENCES

Atkeson, C.-G., Moore, A.-W., and Schaal, S. (1997). Locally weighted learning. *Artif. Intell. Rev.* 11, 11–73. doi: 10.1007/978-94-017-2053-3_2

Belkin, M., and Niyogi, P. (2001). "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proceedings of the 14 th International. Conference on Neural Information Processing Systems*, Cambridge, MA) doi: 10.1162/0899766041732396

Belkin, M., Niyogi, P., and Sindhwani, V. (2006). Manifold regularization: a geometric framework for learning from examples. *J. Mach. Learn. Res.* 7, 2399–2434.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Bottou, L., and Vapnik, V. (1992). Local learning algorithms. *Neur. Comput.* 4, 888–900. doi: 10.1162/neco.1992.4.6.888

Bruzzone, L., and Marconcini, M. (2010). Domain adaptation problems: a DASVM classification technique and a circular validation strategy. *IEEE Trans. Pattern Anal. Mach. Intell* 32, 770–787. doi: 10.1109/TPAMI.2009.57

Chai, X., Wang, Q., Zhao, Y., Liu, X., Bai, O., and Li, Y. (2016). Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition. *Comput. Biol. Med.* 79, 205–214. doi: 10.1016/j.compbiomed.2016.10.019

Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., et al. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* 17:1014.

Chapelle, O. (2006). *Scholkopf Bernhard, and Zien Alexander, Semi-Supervised Learning*. Cambridge, MA: MIT Press.

Chu, W. S., De la Torre, F., and Cohn, J. F. (2017). Selective transfer machine for personalized facial action unit detection. *IEEE Transact. Patt. Anal. Mach. Intellig.* 39, 529–545. doi: 10.1109/CVPR.2013.451

Dan, Y., Tao, J., Fu, J., and Zhou, D. (2021). Possibilistic clustering-promoting semi-supervised learning for EEG-based emotion recognition. *Front. Neurosci.* 15:690044. doi: 10.3389/fnins.2021.690044

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194.

Duan, L., Tsang, I. W., and Xu, D. (2012a). Domain transfer multiple kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 465–479. doi: 10.1109/TPAMI.2011.114

Duan, L., Dong, X., and Shih-Fu, C. (2012b). "Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach," in *Proceedings of the CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, (New York, NY: IEEE), 1338–1345.

Ganin, Y., and Lempitsky, V. (2015). "Unsupervised domain adaptation by back propagation," in *Proceedings of the 32nd International Conference on Machine Learning*, Lille, 1180–1189. doi: 10.1007/s10916-011-9759-1

Gao, J., Wang, S.-T., and Deng, Z.-H. (2010). Global and local preserving based semi-supervised support vector machine. *Acta Electron. Sin.* 38, 1626–1633.

Geng, B., Tao, D., Xu, C., Yang, L., and Hua, X. S. (2012). Ensemble manifold regularization. *IEEE Trans. Pattern Anal. Mach. Intell* 34, 1227–1233. doi: 10.1109/tpami.2012.57

Hou, C., Nie, F., Tao, H., and Yi, D. (2017). Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Trans. Knowl. Data Eng.* 29, 1998–2011. doi: 10.1109/TKDE.2017.2681670

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transact. Affect. Comput.* 5, 327–339. doi: 10.1109/taffc.2014.2339834

Karasuyama, M., and Mamitsuka, H. (2013). Multiple graph label propagation by sparse integration. *IEEE Trans. Neural Netw. Learn. Syst.* 24, 1999–2012. doi: 10.1109/TNNLS.2013.2271327

Kim, M.-K., Kim, M., Oh, E., and Kim, S.-P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Computat. Math. Methods Med.* 2013:573734. doi: 10.1155/2013/573734

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis using physiological signals. *IEEE Transact. Affect. Comput.* 3, 18–31. doi: 10.1109/t-affc.2011.15

Kosko, B. (1986). Fuzzy entropy and conditioning. *Inform. Sci.* 40, 165–174. doi: 10.1016/0020-0255(86)90006-x

Krishnapuram, R., and Keller, J.-M. (1993). A possibilistic approach to clustering [J]. *IEEE Trans. Fuzzy Syst.* 1, 98–110. doi: 10.1109/91.227387

Lan, Z., Sourina, O., Wang, L., Scherer, R., and Müller-Putz, G. R. (2018). Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. *IEEE Transact. Cogn. Dev. Syst.* 11, 85–94. doi: 10.1109/tcds.2018.2826840

Li, Y.-F., and Zhou, Z.-H. (2011). "Improving semi-supervised support vector machines through unlabeled instances selection," in *Proceedings of the 25th AAAI Conference of Artificial. Intelligence*, San Francisco, CA, 386–391. doi: 10.3233/BME-130935

Liu, W., Wang, J., and Chang, S.-F. (2012). Robust and scalable graph-based semi-supervised learning. *Proc. IEEE* 100, 2624–2638.

Long, M., Cao, Y., Wang, J., and Jordan, M. (2015). "Learning transferable features with deep adaptation networks," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Lille, 97–105.

Mansour, Y., Mohri, M., and Rostamizadeh, A. (2009). "Domain adaptation with multiple sources," in *Proceedings of the Conference on Neural Information Processing Systems*, (Vancouver, BC: MIT Press), 1041–1048.

Mühl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain Comput. Interfac.* 1, 66–84. doi: 10.1080/2326263x.2014.912881

Musha, T., Terasaki, Y., Haque, H. A., and Ivamitsky, G. A. (1997). Feature extraction from EEGs associated with emotions. *Artif. Life Robot.* 1, 15–19. doi: 10.1007/bf02471106

Nie, F., Xu, D., Tsang, I., and Zhang, C. (2010). Flexible manifold embedding: a framework for semi-supervised and unsupervised di- mension reduction. *IEEE Transact. Image Proces.* 19, 1921–1932. doi: 10.1109/TIP.2010.2044958

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359.

Shi, X., Guo, Z., Lai, Z., Yang, Y., Bao, Z., Zhang, D., et al. (2015). A framework of joint graph embedding and sparse regression for dimensionality reduction. *IEEE Trans. Image Process.* 24, 1341–1355. doi: 10.1109/TIP.2015.2405474

Tao, J., Dan, Y., and Zhou, D. (2021). Robust multi-source co-adaptation with adaptive loss minimization. *Sign. Proces. Image Commun.* 99:116455. doi: 10.1016/j.image.2021.116455

Tao, J., Wen, S., and Hu, W. (2015). L1-norm locally linear representation regularization multi-source adaptation learning. *Neur. Netw.* 69, 80–98. doi: 10.1016/j.neunet.2015.01.009

Tao, J., Wen, S., and Hu, W. (2016). Multi-source adaptation learning with global and local regularization by exploiting joint kernel sparse representation. *Knowl. Based Syst.* 98, 76–94. doi: 10.1016/j.knosys.2016.01.021

Tao, J., Zhou, D., Liu, F., and Zhu, B. (2019). Latent multi-feature co-regression for visual recognition by discriminatively leveraging multi-source models. *Patt. Recogn.* 87, 296–316. doi: 10.1016/j.neunet.2019.02.007

Tao, J. W., Song, D., Wen, S., and Hu, W. (2017). Robust multi-source adaptation visual classification using supervised low-rank representation. *Patt. Recogn.* 61, 47–65. doi: 10.1016/j.patcog.2016.07.006

Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323. doi: 10.1126/science.290.5500.2319

Tommasi, T., Orabona, F., and Caputo, B. (2014). Learning categories from few examples with multi model knowledge transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 928–941. doi: 10.1109/tpami.2013.197

Tu, W., and Sun, S. (2013). Semi-supervised feature extraction for EEG classification[J]. *Patt. Anal. Appl. Paa* 16, 213–222. doi: 10.1007/s10044-012-0298-2

Wang, Q.-W., Li, Y.-F., and Zhou, Z.-H. (2019). "Partial label learning with unlabeled data," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19)*, Macao, 3755–3761.

Wang, Y., and Chen, S. (2013). Safety-aware semi-supervised classification[J]. *IEEE Tran.s Neur. Netw. Learn Syst.* 24, 1763–1772. doi: 10.1109/tnnls.2013.2263512

Wang, Y.-Y., Chen, S.-C., and Zhou, Z.-H. (2012). New semi-supervised classification method based on modified cluster assumption. *IEEE Transact. Neur. Netw. Learn. Syst.* 23, 689–702. doi: 10.1109/TNNLS.2012.2186825

Wu, M.-S., and Deng, X.-G. (2018). Semi-supervised pattern classification method based on Tri-DE-ELM. *Comp. Eng. Appl.* 54, 109–114.

Xue, H., and Chen, S. (2007). "Alternative robust local embedding," in *Proceedings of International Conference on Wavelet Analysis Pattern Recognition*, Beijing, 591–596.

Xue, H., Chen, S.-C., and Yang, Q. (2011). Structural regularized support vector machine: a framework for structural large margin classifier. *IEEE Transact. Neur. Netw.* 22, 573–587. doi: 10.1109/TNN.2011.2108315

Yang, J., Yan, R., and Hauptmann, A. G. (2007). "Cross-domain video concept detection using adaptive svms," in *Proceedings of the 15th ACM International Conference on Multimedia*, (New York, NY: ACM), 188–197.

Zhang, Y., Dong, J., Zhu, J., and Wu, C. (2019b). Common and special knowledge-driven TSK fuzzy system and its modeling and application for epileptic EEG signals recognition. *IEEE Access* 7, 127600–127614. doi: 10.1109/access.2019.2937657

Zhang, Y., Chung, F. L., and Wang, S. (2019a). Takagi-sugeno-kang. Fuzzy systems with dynamic rule weights. *J. Intell. Fuzzy Syst.* 37, 8535–8550. doi: 10.1016/j.isatra.2017.10.012

Zhang, Y., Li, J., Zhou, X., Zhou, T., Zhang, M., Ren, J., et al. (2019c). A view-reduction based multi-view TSK fuzzy system and its application for textile color classification. *J. Amb. Intellig. Human. Comput.* 9, 1–11. doi: 10.1007/s12652-019-01495-9

Zhang, Y., Tian, F., Wu, H., Geng, X., Qian, D., Dong, J., et al. (2017). Brain MRI tissue classification based fuzzy clustering with competitive learning. *J. Med. Imag. Health Inform.* 7, 1654–1659. doi: 10.1006/cbmr.1996.0023

Zhang, Y., Wang, L., Wu, H., Geng, X., Yao, D., and Dong, J. (2016). A clustering method based on fast exemplar finding and its application on brain magnetic resonance images segmentation [J]. *J. Med. Imag. Health Inform.* 6, 1337–1344. doi: 10.1166/jmihi.2016.1923

Zhang, Y., Wang, S., Xia, K., Jiang, Y., and Qian, P. (2020). Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inform. Fus.* 66, 170–183. doi: 10.1016/j.inffus.2020.09.002

Zhao, Y., Liebgott, H., and Cachard, C. (2015). Comparison of the existing tool localization methods on two-dimensional ultrasound images and their tracking results. *Control Theor Appl.* 9, 1124–1134. doi: 10.1049/iet-cta.2014.0672

Zhao, Y., Shen, Y., Bernard, A., Cachard, C., and Liebgott, H. (2016). Evaluation and comparison of current biopsy needle localization and tracking methods using 3d ultrasound. *Ultrasonics* 73, 206–220. doi: 10.1016/j.ultras.2016.09.006

Zheng, W.-L., and Lu, B.-L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transact. Autono. Mental Dev.* 7, 162–175. doi: 10.1109/tamd.2015.2431497

Zhong, P., Wang, D., and Miao, C. (2020). EEG-based emotion recognition using regularized graph neural networks. *IEEE Transact. Affect. Comput.* 99:1. doi: 10.1109/taffc.2020.2994159

Zhou, S.-H., Liu, X.-W., Zhu, C.-Z., Liu, Q., and Yin, J. (2014). "Spectral clustering-based local and global structure preservation for feature selection," in *Proceedings of the 2014 International Joint Conference on Neural Networks*, Beijing, 550–557.

Zhou, X., Jin, K., Shang, Y., and Guo, G. (2018). Visually interpretable representation learning for depression recognition from facial Im-ages. *IEEE Transact. Affect. Comput.* 11, 542–552. doi: 10.1109/TAFFC.2018.2828819

Zhu, X.-J. (2008). *Semi-Supervised Learning Literature Survey [R]. Computer Science TR 1530*. Madison, WI: University of Wis-consin.

Zhu, X.-J., and Goldberg, A. (2009). *Introduction to Semi-Supervised Learning*, Vol. 3. San Rafael, CA: Morgan & Claypool, 130.

Zhu, Y., Shang, Y., Shao, Z., and Guo, G. (2017). automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Transact. Affect. Comput.* 9, 578–584. doi: 10.1109/TAFFC.2017.2650899

Zu, B.-B., Xia, K.-W., Wen, L., Niu, W.-J., and Jiang, X.-Q. (2019). Semi-supervised classification application of remote sensing image based on block low rank images. *J. Front. Comput. Sci. Technol.* 13, 1217–1226.

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 Dan, Tao and Zhou. 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 Neuroscience

ORIGINAL RESEARCH
published: 10 June 2022
doi: 10.3389/fnins.2022.837041

A Radiomics Nomogram for Classifying Hematoma Entities in Acute Spontaneous Intracerebral Hemorrhage on Non-contrast-Enhanced Computed Tomography

Jia Wang[1†], Xing Xiong[2†], Jing Ye[1], Yang Yang[1], Jie He[1], Juan Liu[1] and Yi-Li Yin[1]*

[1] Department of Radiology, Northern Jiangsu People's Hospital, Yangzhou, China, [2] Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China

OPEN ACCESS

Edited by:
Yuanpeng Zhang,
Nantong University, China

Reviewed by:
Nianyin Zeng,
Xiamen University, China
Ying Li,
Fudan University, China

***Correspondence:**
Yi-Li Yin
sbrmyinyili@163.com

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

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 16 December 2021
Accepted: 02 May 2022
Published: 10 June 2022

Citation:
Wang J, Xiong X, Ye J, Yang Y, He J, Liu J and Yin Y-L (2022) A Radiomics Nomogram for Classifying Hematoma Entities in Acute Spontaneous Intracerebral Hemorrhage on Non-contrast-Enhanced Computed Tomography. Front. Neurosci. 16:837041. doi: 10.3389/fnins.2022.837041

Aim: To develop and validate a radiomics nomogram on non-contrast-enhanced computed tomography (NECT) for classifying hematoma entities in patients with acute spontaneous intracerebral hemorrhage (ICH).

Materials and Methods: One hundred and thirty-five patients with acute intraparenchymal hematomas and baseline NECT scans were retrospectively analyzed, i.e., 52 patients with vascular malformation-related hemorrhage (VMH) and 83 patients with primary intracerebral hemorrhage (PICH). The patients were divided into training and validation cohorts in a 7:3 ratio with a random seed. After extracting the radiomics features of hematomas from baseline NECT, the least absolute shrinkage and selection operator (LASSO) regression was applied to select features and construct the radiomics signature. Multivariate logistic regression analysis was used to determine the independent clinical-radiological risk factors, and a clinical model was constructed. A predictive radiomics nomogram was generated by incorporating radiomics signature and clinical-radiological risk factors. Nomogram performance was assessed in the training cohort and tested in the validation cohort. The capability of models was compared by calibration, discrimination, and clinical benefit.

Results: Six features were selected to establish radiomics signature *via* LASSO regression. The clinical model was constructed with the combination of age [odds ratio (OR): 6.731; 95% confidence interval (CI): 2.209–20.508] and hemorrhage location (OR: 0.089; 95% CI: 0.028–0.281). Radiomics nomogram [area under the curve (AUC), 0.912 and 0.919] that incorporated age, location, and radiomics signature outperformed the clinical model (AUC, 0.816 and 0.779) and signature (AUC, 0.857 and 0.810) in the

training cohort and validation cohorts, respectively. Good calibration and clinical benefit of nomogram were achieved in the training and validation cohorts.

Conclusion: Non-contrast-enhanced computed tomography-based radiomics nomogram can predict the individualized risk of VMH in patients with acute ICH.

Keywords: radiomics, nomogram, non-contrast-enhanced computed tomography, intracerebral hemorrhage, vascular malformations-related hemorrhage

INTRODUCTION

Spontaneous intracerebral hemorrhage (ICH) is defined as intraparenchymal hemorrhage caused by vascular rupture, which can be divided into primary hemorrhage and secondary hemorrhage (Meretoja et al., 2012). Primary intracerebral hemorrhage (PICH) is mainly due to spontaneous rupture of blood vessels caused by hypertension, arteriosclerosis, or amyloidosis, while secondary cerebral hemorrhage generally refers to the existence of secondary cause, such as the hemorrhagic transformation of ischemic stroke, cerebral venous thrombosis, tumors, and vascular malformation-related hemorrhage (VMH). For patients suspected of acute ICH, non-contrast-enhanced computed tomography (NECT) scan remains the first choice of imaging during a clinical emergency, which can effectively reflect the location, volume, and surrounding brain tissue structure of hemorrhage (Thabet et al., 2017). Several studies have shown a major difference in treatment and prognosis between VMH and PICH (Barone et al., 2017; Fukuda et al., 2017). For PICH, the primary strategy is to monitor vital signs and hemorrhage changes, stabilize blood pressure, and prevent further bleeding. Unless the patient is suffering from severe intracranial hypertension or cerebral hernia, surgical treatment should not be considered. Ruptured VMH can easily re-hemorrhage, so surgical resection or interventional embolization has a positive effect on the prognosis of patients.

It is crucial to distinguish VMH from PICH in the early stage. However, it is difficult and equivocal to distinguish hemorrhage types by the naked eye on NECT images. As a quantitative imaging research method, radiomics can non-invasively evaluate regional tissue heterogeneity at the millimeter scale, which has been confirmed in tumor imaging (Lambin et al., 2017). However, the radiomics research in hemorrhagic diseases is limited, and no study has reported a NECT-based radiomics nomogram to differentiate VMH from PICH.

In this study, we hypothesized that VMH contains unique NECT-based radiomics features, and we aimed to develop and validate a radiomics nomogram for the classification of VMH, to facilitate individualized treatment strategies for patients with acute ICH.

MATERIALS AND METHODS

Subjects

This study was approved by the Ethics Committee of the local institution, and informed consent was waived off. The patients were retrospectively analyzed between October 2017 and September 2020, i.e., 333 patients with acute ICH who received baseline NECT scan after symptom onset. The inclusion criteria were as follows: (1) acute, non-traumatic intraparenchymal hemorrhage; (2) NECT scan within 6 h of the onset of ICH symptoms; and (3) complete medical records. Patients with the following conditions were excluded: (1) acute brain injury ($n = 56$); (2) isolated intraventricular hemorrhage, subarachnoid hemorrhage, epidural hemorrhage, or venous sinus embolism ($n = 43$); (3) hemorrhage related to tumor, aneurysm, or ischemic stroke ($n = 36$); (4) multiple hemorrhages ($n = 9$); (5) baseline NECT scan after 6 h of the onset ($n = 39$); and (6) serious image artifacts ($n = 15$).

A total of 135 consecutive patients (95 men and 40 women; mean age, 54.22 $\pm$ 15.38 years) were enrolled, and their demographic information was collected from medical records. VMH was diagnosed according to post-digital subtraction angiography (DSA) and/or surgery and repeated DSAs were operated if considered necessary. Among them, 41 were confirmed as arteriovenous malformations, 7 as cavernous malformations, and 4 as arteriovenous fistula. The database was divided into a training cohort ($n = 95$; 65 men and 30 women; mean age, 54.06 $\pm$ 15.70 years) and a validation cohort ($n = 40$; 30 men and 10 women; mean age, 54.60 $\pm$ 14.61 years) in a 7:3 ratio with a random seed. Models were developed in the training cohort and independently tested in the validation cohort.

CT Imaging

All patients underwent baseline head NECT scan using a GE Gemstone Scanner. The scan protocol was as follows: (1) tube voltage, 120 kV; (2) tube current, 360 mA; (3) field of view, 320 mm; (4) matrix, 512 $\times$ 512; and (5) slice thickness, 5 mm.

Qualitative Analysis of Non-contrast-Enhanced Computed Tomography Images

All CT images were analyzed by two radiologists with one having experience of 7 years and another with 5 years independently. They were blinded to the clinical information. The two radiologists evaluated the following image features of hemorrhage: (1) location (deep/lobar) and (2) shape (regular or irregular). A discussion was carried out to achieve consensus in case of discrepancy.

Clinical-Radiological Risk Factors

The univariate analysis was used to assess the single clinical and NECT features for discriminating VMH from PICH in the training cohort, and a multivariate logistic regression analysis

A $z\text{-score} = (x - \mu)/\sigma$

B $NB = (TP/N - FP/N) * Pt/(1-Pt)$

C $NRI = (C1-B1)/N1 + (B2-C2)/N2$

D $IDI = (\bar{P}new, events - \bar{P}old, events) - (\bar{P}new, nonevents - \bar{P}old, nonevents)$

FIGURE 1 | Formulas of *z*-score **(A)**, net benefit **(B)**, net reclassification index **(C)**, and integrated discrimination index **(D)**. *x*, the feature value; μ, the average of the feature values among all VOIs; σ, the corresponding standard deviation; NB, net benefit; TP, true positive; FP, false positive; *N*, total cases; Pt, probability threshold; NRI, net reclassification index; C1, B1, the cases predicted by the new and old model in the event occurrence group; B2, C2, the cases predicted by the new and old model in the event incurrence group; N1, N2, the cases in the event occurrence and incurrence group; IDI, integrated discrimination index; −Pnew, events,−Pold, events, the mean of the probability of disease predicted by the new and old model for each individual in the event occurrence group; Pnew, non-events,−Pold, non-events, the mean of the probability of disease predicted by the new and old model for each individual in the event incurrence group.

with significant variables was performed to determine the potential risk factors of VMH.

Image Processing and Segmentation

The voxel-based segmentation of hemorrhage was performed using 3D-Slicer software (version 3.6.0[1]) by one radiologist. Regions of interest (ROI) were semi-automatically delineated on each slice of the CT image containing the entire lesion. The segmentations were then validated by another radiologist in a cohort of 30 randomly selected patients. All images were resampled to 1 mm × 1 mm × 1 mm voxel and normalized by histogram matching to eliminate the intensity difference.

Extraction of Radiomics Feature

Prior to extracting the features, image filtration was implemented on the original image with wavelet (wavelet-LHL, wavelet-LHH, wavelet-HLL, wavelet-LLH, wavelet-HLH, wavelet-HHH, wavelet-HHL, and wavelet-LLL) and Laplacian of Gaussian (LoG) transform. The log filtration was calculated with operator values of 1 mm (fine), 3 mm (medium), and 5 mm (coarse), respectively. A set of 1,130 radiomics features was extracted from each original and filtered segmentation and divided into five groups: (1) intensity (histogram-derived first-order statistics (n = 18); (2) shape (n = 14); (3) textural matrix [i.e., the gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), and the neighborhood gray-tone difference matrix (NGTDM), n = 75]; (4) wavelet-based transform (n = 744), and (5) log-based transform (n = 279). All features were automatically extracted from the volumes of interest (VOIs), which contain 3D information on hematoma using PyRadiomics, and the detailed descriptions of radiomics features could be found on the https://pyradiomics.readthedocs.io/en/latest/index.html.

Selection of Radiomics Feature

Inter-class correlation (ICC) coefficient was calculated for each radiomics feature, and a single-factor logistic regression analysis was carried out to select highly significant and correlated features. Only the features with ICC > 0.8 and $p < 0.1$ were considered as high repeatability and correlated, which were included for subsequent analysis. Then, the feature values of all VOIs were normalized with *z*-score normalization (**Figure 1A**) to eliminate the unit limit for each feature. The least absolute contraction and selection operator (LASSO) was applied to select features to reduce the redundancy and overfitting in the training cohort. The set of non-zero coefficient features was chosen with optimal λ value, which was determined by the area under the curve (AUC) of fivefold cross-validation. The radiomics score (Rad-score) of each patient was calculated using the linear fitting of selected features, and the radiomics signature was constructed. The formula was as follows: $Rad - score = \alpha + \sum_{1}^{i} \beta_i X_i$ (α, the intercept; β_i, the value of radiomics feature selected by LASSO; β_i, the corresponding coefficients).

Construction and Evaluation of the Prediction Model

After the selection of potential clinical-radiological risk factors of VMH based on the multivariate logistic regression analysis, the clinical model in the training cohort was built. Then, the radiomics nomogram model incorporating the clinical-radiological risk factors and radiomics signature were also constructed. The calibration curves and the Hosmer–Lemeshow test were performed to evaluate the goodness-of-fit of the nomogram. Receiver operating characteristic (ROC) curve analysis was used to evaluate the discrimination performance of the models. Decision curve analysis (DCA) was performed to determine the clinical benefit of the models by calculating their net benefits (**Figure 1B**). The categorical net reclassification index (NRI, three risk categories: <0.5, 0.5–0.8, and >0.8, **Figure 1C**) and total integrated discrimination index (IDI, **Figure 1D**) were calculated to evaluate the prediction improvement of the radiomics nomogram when compared to clinical-radiological factors.

Statistical Analysis

Student's *t*-test, the Mann–Whitney *U* test, the Chi-square test, and Fisher's exact test were used for univariate analysis, as appropriate. The ROC curves were compared using the DeLong test. The statistical analysis was performed using R (version 3.6.2, Boston, MA, United States) and SPSS software (version 22,

[1]http://www.slicer.org

Chicago, IL, United States). Two-tailed values of $p < 0.05$ were considered as statistically significant.

RESULTS

Patient Characteristics

There were significant differences in age, history of hypertension, and location between the VMH group and the PICH group in the training cohorts ($p < 0.001$), but there were no significant differences in gender, history of smoking, diabetes, alcohol consumption, and coagulopathy or shape ($p > 0.05$; **Table 1**). No differences were found in clinical-radiological characteristics between the training and validation cohorts ($p = 0.137$–0.922).

Clinical Model Construction

Multivariable analysis showed that age (subdivided into ≥50 and <50 years old classes, $p = 0.001$) and location ($p < 0.001$) were independent risk factors for clinical model construction. Patients with younger age [odds ratio (OR): 6.731; 95% confidence interval (CI): 2.209–20.508] or hemorrhages located in the lobar (OR: 0.089; 95% CI: 0.028–0.281) were likely to be VMH.

Radiomics Signature Construction

Of the 1,130 radiomics features extracted from NECT images, 532 features were demonstrated with ICC > 0.8 and $p < 0.1$. Six VMH-related features with non-zero coefficients were selected using a LASSO logistic regression for Rad-score calculation, and a radiomics signature was constructed (**Figures 2A,B**). The selected radiomics features and the corresponding coefficients are displayed in **Table 2**. The Rad-score differed significantly between the VMH and PICH groups in both training ($p < 0.001$) and validation ($p < 0.05$) cohorts (**Figures 2C,D**). Patients with VMH generally had lower Rad-scores than those with PICH.

The Radiomics Nomogram Development and Model Assessment

Age, location, and radiomics signature were incorporated into the development of radiomics nomogram (**Figure 3A**). Patients with younger age (<50 years) or hemorrhages located in lobes and the negative radiomics signature were more likely to be VMH. **Figure 3B** shows that the calibration curve of the nomogram fitted well with the actual status. The Hosmer–Lemeshow test showed good calibration in the training cohort ($p = 0.914$).

The diagnostic performance of each model is summarized in **Table 3**, and the ROC curves are shown in **Figure 4**. Compared with radiomics signature (AUC: 0.857; 95% CI: 0.771–0.921; sensitivity, 0.763; and specificity, 0.889) and clinical model (AUC: 0.816; 95% CI: 0.723–0.888; sensitivity, 0.661; and specificity, 0.806), the radiomics nomogram (AUC: 0.912; 95% CI: 0.836–0.960; sensitivity, 0.746; and specificity, 0.889) showed better performance in differentiating VMH from PICH. The DeLong test showed that the performance of the radiomics nomogram had a statistically significant improvement ($p < 0.05$).

Model Validation

Figure 3C shows the calibration curve of the nomogram in the validation cohort with a non-significant Hosmer–Lemeshow test statistic ($p = 0.720$). The AUC of the radiomics nomogram (AUC: 0.919; 95% CI: 0.788–0.982) was higher than that of the clinical model (AUC: 0.779; 95% CI: 0.655–0.916) and radiomics signature (AUC: 0.810; 95% CI: 0.788–0.982) in the validation cohort.

The DCA for the three models is shown in **Figure 5**. The curve showed that the radiomics nomogram had a higher overall net benefit in differentiating VMH from PICH than the clinical model and radiomics signature, within the threshold probability <0.8.

The discrimination measures confirmed that adding radiomics signature to clinical-radiological factors significantly improved reclassification for predicting hemorrhage status, with NRI and IDI of 0.435 (95% CI: 0.178–0.692, $p < 0.001$) and 0.170 (95% CI: 0.092–0.247, $p < 0.001$), 0.313 (95% CI: −0.049 to 0.674, $p < 0.001$), and 0.250 (95% CI: 0.103–0.3961, $p < 0.001$) in the training and validation cohorts, respectively.

DISCUSSION

In this study, we constructed and evaluated three predictive models to non-invasively distinguish VMH from PICH, such

TABLE 1 | Demographic data and NECT features.

Variables	Training cohort ($n = 95$)			Validation cohort ($n = 40$)		
	VMH	PICH	*p*	VMH	PICH	*p*
Gender (male/female)	26/10	39/20	0.534	11/5	19/5	0.456
Age (years, mean ± SD)	42.36 ± 17.03	61.20 ± 14.89	<0.001	41.75 ± 15.70	63.17 ± 13.88	<0.001
Hypertension (P/N)	13/23	42/17	<0.001	2/14	16/8	0.001
Diabetes (P/N)	3/33	10/49	0.358	1/15	3/21	0.638
Smoking (P/N)	2/34	11/48	0.122	3/13	6/18	0.717
Alcohol consumption (P/N)	1/35	7/52	0.252	1/15	4/20	0.631
Coagulopathy (P/N)	0/36	3/56	0.286	0/16	3/21	0.262
Location (deep/lobar)	7/29	39/20	<0.001	5/11	14/10	0.093
Shape (regular/irregular)	20/16	36/23	0.600	5/11	13/11	0.154

NECT, non-contrast-enhanced CT; VMH, vascular malformations-related hemorrhage; PICH, primary intracerebral hemorrhage; P, positive; N, negative.

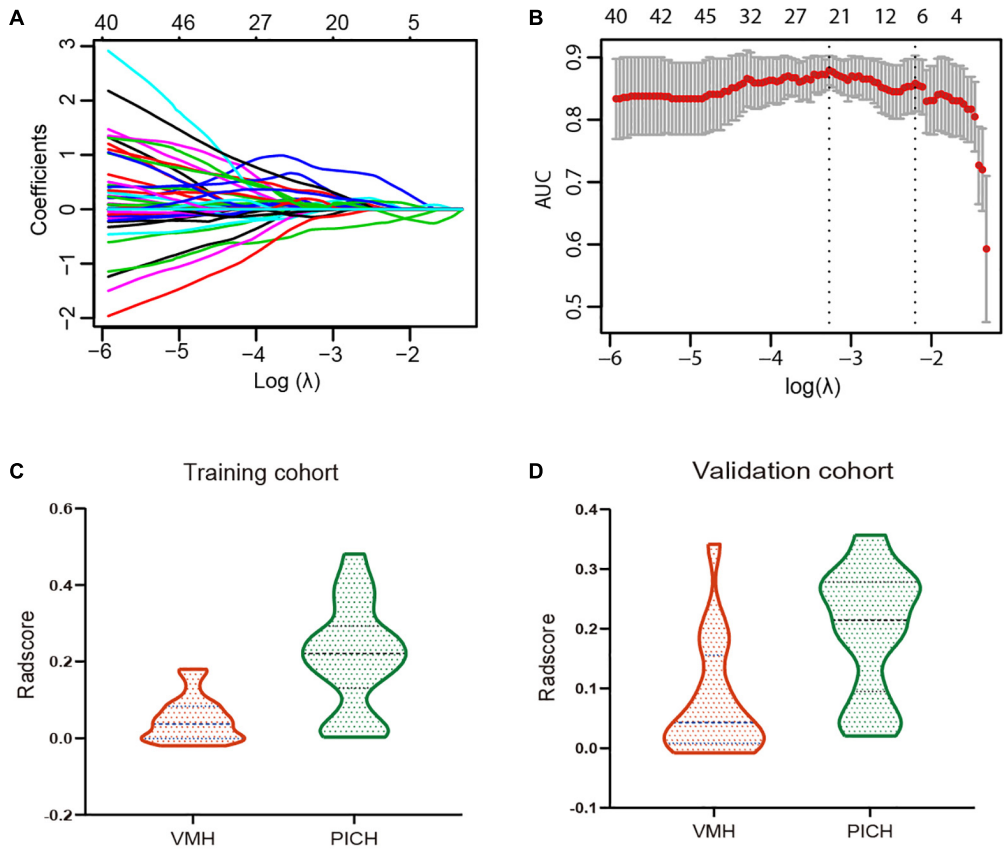


FIGURE 2 | (A) Tuning parameter (λ) selection in the least absolute shrinkage and selection operator (LASSO) regression used fivefold cross-validation *via* one standard error (SE) of the minimum criteria. The optimal λ value was 0.1212 with log(λ) = 2.1095. **(B)** LASSO regression coefficient of the 532 radiomics features generated vs. the selected log(λ) value. **(C,D)** The violin plots for radiomics signature in the training and validation cohorts, categorized by the vascular malformations-related hemorrhage (VMH) and PICH groups.

as the clinical model, radiomics signature, and radiomics nomogram. The radiomics nomogram incorporated age, hemorrhage location, and radiomics signature derived from NECT and showed satisfactory performance for predicting VMH. Thus, it provides a straightforward and safe auxiliary approach for a clinical emergency.

Patients with VMH-related hemorrhage have lower inpatient mortality and are more likely to have a favorable discharge disposition as compared to patients with PICH (Murthy et al., 2017). The timely surgical or interventional practice has a positive impact on the outcomes of patients with VMH, otherwise, the effect worsens after 48 h of the onset of symptoms. Therefore, patients with VMH should be screened as soon as possible for further operation, instead of receiving conservative treatment as the patients with PICH. The PICH is diagnosed by excluding a thorough investigation for secondary structural causes of ICH. Older age, deep location, and history of hypertension are usually considered a diagnostic basis of PICH, although cerebral angiography studies suggest that these imaging and clinical features are not always reliable indicators in selecting patients for further therapy, and patients with these features may have coexisting vascular abnormalities (Josephson et al., 2014; van Asch et al., 2015).

Prior to radiomics analysis, we evaluated the clinical-radiological factors. In our study, age and hemorrhage location were independent risk factors for distinguishing PICH and VMH, similar to the findings of Murthy et al. (2017). Several studies have shown that when compared with PICH, the VMH tended to be more irregular, and the bottom of the lesion was usually located on the cortex, with the tip of the lesion pointed to the lateral ventricle (Wagle et al., 1984). Unexpectedly, in our

TABLE 2 | The selected radiomics features and coefficients.

Image	Class	Name	Coefficients
Sigma-5 mm	GLCM	Lmc2	−0.1157
Sigma-5 mm	GLCM	Lmc1	0.4231
Wavelet-HLL	GLCM	Idn	−0.2140
Wavelet-LHL	GLCM	Correlation	0.0068
Wavelet-LLL	GLSZM	GrayLevelNonUniformityNormalized	0.1238
Original	GLSZM	LowGrayLevelZoneEmphasis	0.0331

GLCM, gray level co-occurrence matrix; GLSZM, gray level size zone matrix.

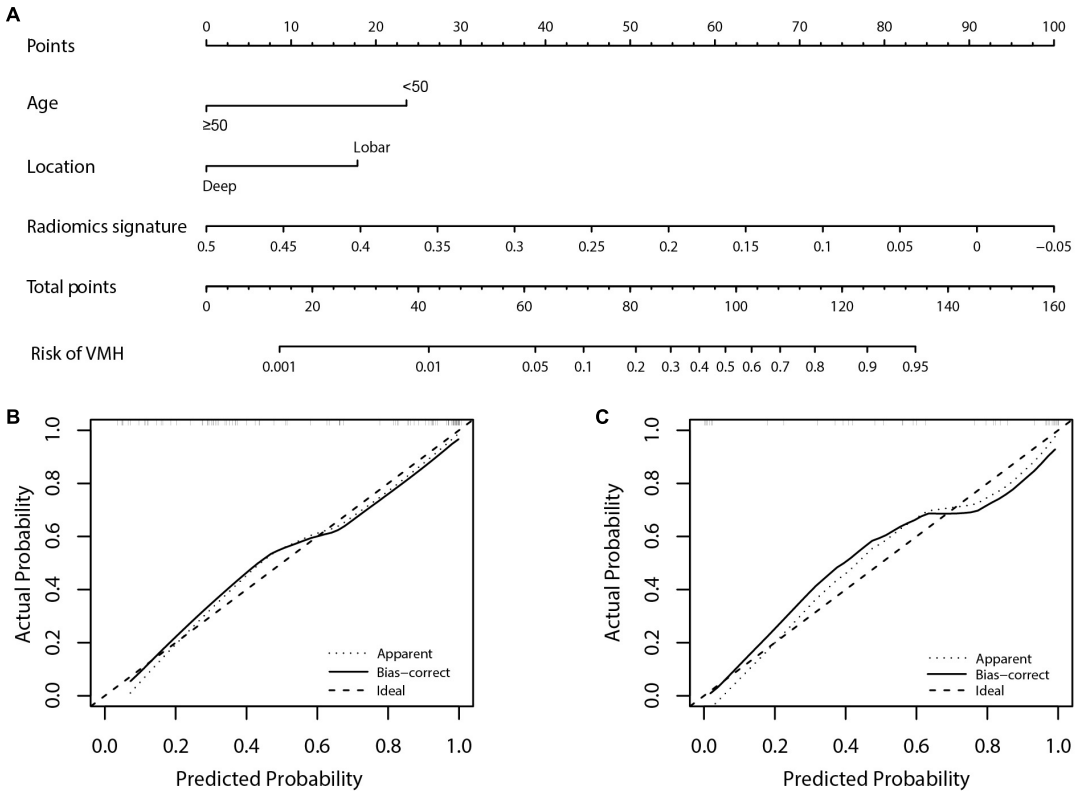


FIGURE 3 | The radiomics nomogram for predicting the status of hemorrhage and calibration curves. **(A)** The radiomics nomogram, combining age, location, and radiomics signature were developed in the training cohort. Patients with younger age (<50 years) or hemorrhages located in lobes and the negative radiomics signature were more likely to be vascular malformations-related hemorrhage (VMH). **(B,C)** Calibration curves for the radiomics nomogram in the training and validation cohorts. Calibration curves indicated the agreement of the nomogram between the predicted risk of VMH and actual status. The 45° gray dotted line represents the ideal prediction and the black curve is the predictive performance of the nomogram. The closer the black curve is to the gray dotted line, the better is the predictive efficacy of the nomogram.

TABLE 3 | Predictive performance of radiomics nomogram, radiomics signature, and clinical model.

Model	Training cohort				Validation cohort			
	AUC (95% CI)	SEN	SPE	Youden index	AUC (95% CI)	SEN	SPE	Youden index
Clinical model	0.816 (0.723–0.888)	0.661	0.806	0.467	0.779 (0.620–0.894)	0.833	0.688	0.521
Radiomics signature	0.857 (0.771–0.921)	0.763	0.889	0.652	0.810 (0.655–0.916)	0.667	0.875	0.542
Radiomics nomogram	0.912 (0.836–0.960)	0.746	0.889	0.635	0.919 (0.788–0.982)	0.958	0.813	0.771

AUC, area under the curve; CI, confidence interval; SEN, sensitivity; SPE, specificity.

study, the shape of the hemorrhage was failed to show enough predictive strength based on univariable association with VMH. This may be due to the evaluation of shape (regular/irregular) being influenced by the reader's academic background and experience. Moreover, the descriptors of two-/three-dimensional size and shape of lesions were included in radiomics features. For example, sphericity is a measure of the roundness of the shape of the region relative to a sphere (Lorensen et al., 2020). Hence, the integration of shape described by radiomics features can decrease bias when compared with visual evaluation in theory, which makes the exclusion of this variable an appropriate strategy for model development.

Radiomics, using different machine learning methods to construct predictive models, can non-invasively reflect the internal heterogeneity of lesions to achieve early diagnosis, differential diagnosis, and treatment response monitoring. This method has been widely studied in various tumor-related research studies (Kocak et al., 2018; Artzi et al., 2019; Shu et al., 2019). Compared with PICH, the internal composition and morphology of VMH can be more complex and heterogeneous due to the existence of malformed blood vessels, which can be reflected by radiomics features. So far, radiomics analysis of vascular diseases has mainly focused on the identification and stratification of stability of vessel plaque (Shi et al., 2018;

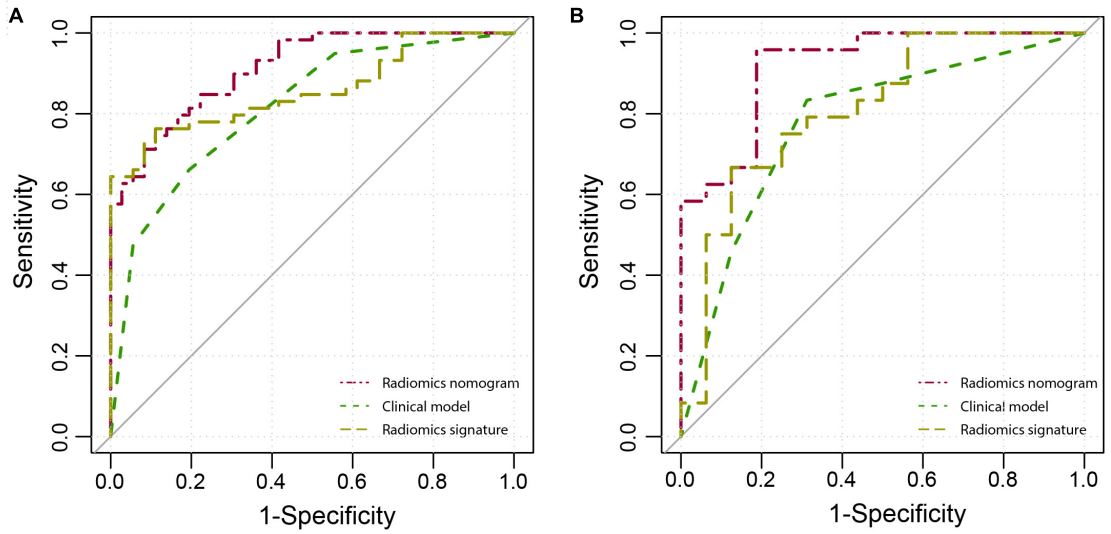


FIGURE 4 | The receiver operator characteristic (ROC) curves of the radiomics nomogram, clinical model, and radiomics signature in the training **(A)** and validation **(B)** cohorts.

Kolossvary et al., 2019), the prediction of cerebral hemorrhage expansion (Ma et al., 2019; Xie et al., 2020; Xu et al., 2020), and discrimination of tumorous ICH from benign causes (Choi et al., 2015; Nawabi et al., 2020). The current studies showed that radiomics features may have the advantages of a higher resolution and objectively quantify the heterogeneity of hematoma. In our study, six radiomics features belonging to the GLSZM and GLCM that described the heterogeneity of the hematoma may be associated with VMH.

To the best of our knowledge, only one reported study applied radiomics analysis to distinguish the hemorrhage caused by arteriovenous malformation and other etiologies (Zhang et al., 2019). Zhang et al. (2019) constructed 88 predictive models by the combination of 11 feature-selection methods and 8 classifiers predictive models, and then the optimal one was selected and evaluated. However, they included fewer radiomics features. Moreover, CT imaging features and clinical characteristics were excluded from their study. In our study, we extracted and analyzed 1,130 radiomics features, which can more comprehensively describe the internal heterogeneity of hemorrhage. After processing, five of the six optimal features were derived from LoG and wavelet filtration, which cannot be obtained by conventional texture analysis. As a result, with the aid ofmodeling NECT-based radiomics, the post-processing and reporting time are limited to 5 min. More importantly, the radiomics nomogram can generate an individual probability of VMH with its visual scoring system, which meets the requirement of personalized medicine. Additionally, DCA, NRI, and IDI analyses were applied in the current study to assess the clinical improvement of radiomics nomogram-assisted decisions on patient outcomes. The DCA showed that using the radiomics nomogram to predict VMH obtains more benefits than either the treat-all-patients or the treat-none scheme. The NRI and IDI analyses confirmed the reclassification improvement by adding radiomics signature to clinical-radiological factors.

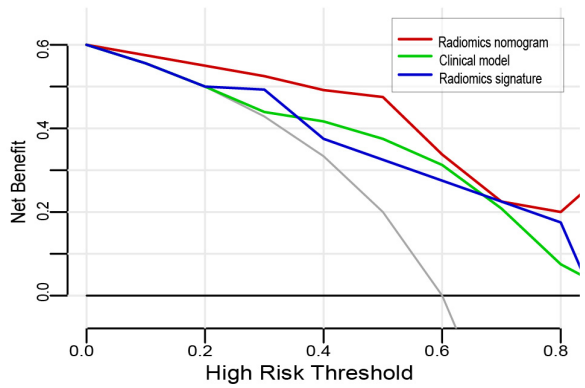


FIGURE 5 | Decision curve analysis for the three models. The *y*-axis indicates the net benefit; the *x*-axis indicates the threshold probability. The red line, green line, and blue line represent the net benefit of the radiomics nomogram, clinical model, and radiomics signature, respectively. The radiomics nomogram had the highest net benefit compared with the other two models when the threshold probability was <0.8 at which a patient would be diagnosed as vascular malformations-related hemorrhage (VMH).

The ability of radiomics model to distinguish different hematoma types in a reliable and reproducible manner represents an attractive alternative in centers without ready access to either advanced imaging modalities or stroke neurologists and/or neuroradiologists for imaging interpretation. Additionally, this approach offers other potential benefits that include (1) reduced contrast and radiation exposure and (2) faster times to treatment, which theoretically would lead to improved outcomes. Finally, this tool might become complementary rather than competing approaches. There were some limitations in this study. First was

the retrospective nature and small sample size of this study. Second, our analysis lacked independent external validation. Hence, prospective studies are required with larger cohorts from external multicenter validation. Last, the semiautomatic methods to segment hemorrhage may be time-consuming. Future studies could explore the role of artificial intelligence in providing a rapid comprehensive segmentation (Zeng et al., 2018, 2021, 2022; Wu et al., 2022).

In summary, the NECT-based radiomics nomogram was developed and validated in this study, which provides a valuable tool for the individualized risk prediction of VMH in patients with ICH. As this information is not assessable by human eyes, the proposed approach can be used as a supportive tool to improve the radiologist's diagnostic decision. Radiomics nomogram, as a non-invasive and quantitative method, may serve as a promising tool to complement the conventional procedures for the clinical decision-making process in future large-scale applications.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Northern Jiangsu 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

Y-LY and JY guaranteed the manuscript. JW and XX contributed to conception, design, collection, and assembly of data. YY, JL, and JH involved in data analysis and interpretation. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Social Develop Foundation of Yangzhou (no. 2017066), Yangzhou City Science and Education Strengthening Leading Talents Project (no. LJRC201810), Yangzhou City Science and Education Strengthening Key Talents Project (no. ZDRC201873), and Jiangsu Province "Six First Project" for High-Level Health Professionals (no. LGY2019032).

REFERENCES

Artzi, M., Bressler, I., and Ben, B. D. (2019). Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *J. Magn. Reson. Imaging* 50, 519–528. doi: 10.1002/jmri.26643

Barone, D. G., Marcus, H. J., Guilfoyle, M. R., Higgins, J. N. P., Antoun, N., Santarius, T., et al. (2017). Clinical Experience and Results of Microsurgical Resection of Arterioveonous Malformation in the Presence of Space-Occupying Intracerebral Hematoma. *Neurosurgery* 81, 75–86. doi: 10.1093/neuros/nyx003

Choi, Y. S., Rim, T. H., Ahn, S. S., and Lee, S. K. (2015). Discrimination of Tumorous Intracerebral Hemorrhage from Benign Causes Using CT Densitometry. *AJNR Am. J. Neuroradiol.* 36, 886–892. doi: 10.3174/ajnr.A4233

Fukuda, K., Majumdar, M., Masoud, H., Nguyen, T., Honarmand, A., Shaibani, A., et al. (2017). Multicenter assessment of morbidity associated with cerebral arteriovenous malformation hemorrhages. *J. Neurointerv. Surg.* 9, 664–668. doi: 10.1136/neurintsurg-2016-012485

Josephson, C. B., White, P. M., Krishan, A., and Al-Shahi Salman, R. (2014). Computed tomography angiography or magnetic resonance angiography for detection of intracranial vascular malformations in patients with intracerebral haemorrhage. *Cochrane Database Syst. Rev.* 2014:CD009372. doi: 10.1002/14651858.CD009372

Kocak, B., Yardimci, A. H., Bektas, C. T., Turkcanoglu, M. H., Erdim, C., Yucetas, U., et al. (2018). Textural differences between renal cell carcinoma subtypes: machine learning-based quantitative computed tomography texture analysis with independent external validation. *Eur. J. Radiol.* 107, 149–157. doi: 10.1016/j.ejrad.2018.08.014

Kolossvary, M., Karady, J., Kikuchi, Y., Ivanov, A., Schlett, C. L., and Lu, M. T. (2019). Radiomics versus Visual and Histogram-based Assessment to Identify Atheromatous Lesions at Coronary CT Angiography: an ex Vivo Study. *Radiology* 293, 89–96. doi: 10.1148/radiol.2019190407

Lambin, P., Leijenaar, R., 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, 749–762. doi: 10.1038/nrclinonc.2017.141

Lorensen, W. E., Johnson, C., Kasik, D., and Whitton, M. C. (2020). History of the Marching Cubes Algorithm. *IEEE Comput. Graph. Appl.* 40, 8–15. doi: 10.1109/MCG.2020.2971284

Ma, C., Zhang, Y., Niyazi, T., Wei, J., Guocai, G., and Liu, J. (2019). Radiomics for predicting hematoma expansion in patients with hypertensive intraparenchymal hematomas. *Eur. J. Radiol.* 115, 10–15. doi: 10.1016/j.ejrad.2019.04.001

Meretoja, A., Strbian, D., Putaala, J., Curtze, S., Haapaniemi, E., and Mustanoja, S. (2012). SMASH-U: a proposal for etiologic classification of intracerebral hemorrhage. *Stroke* 43, 2592–2597. doi: 10.1161/STROKEAHA.112.661603

Murthy, S. B., Merkler, A. E., Omran, S. S., Gialdini, G., Gusdon, A., Hartley, B., et al. (2017). Outcomes after intracerebral hemorrhage from arteriovenous malformations. *Neurology* 88, 1882–1888. doi: 10.1212/WNL.0000000000003935

Nawabi, J., Kniep, H., Kabiri, R., Broocks, G., Faizy, T. D., Thaler, C., et al. (2020). Neoplastic and Non-neoplastic Acute Intracerebral Hemorrhage in CT Brain Scans: machine Learning-Based Prediction Using Radiomic Image Features. *Front. Neurol.* 11:285. doi: 10.3389/fneur.2020.00285

Shi, Z., Zhu, C., Degnan, A. J., Tian, X., Li, J., Chen, L., et al. (2018). Identification of high-risk plaque features in intracranial atherosclerosis: initial experience using a radiomic approach. *Eur. Radiol.* 28, 3912–3921. doi: 10.1007/s00330-018-5395-1

Shu, Z., Fang, S., Ding, Z., Mao, D., Cai, R., Chen, Y., et al. (2019). MRI-based Radiomics nomogram to detect primary rectal cancer with synchronous liver metastases. *Sci. Rep.* 9:3374. doi: 10.1038/s41598-019-39651-y

Thabet, A. M., Kottapally, M., and Hemphill, J. C. III (2017). Management of intracerebral hemorrhage. *Handb. Clin. Neurol.* 140, 177–194. doi: 10.1016/B978-0-444-63600-3.00011-8

van Asch, C. J., Velthuis, B. K., Rinkel, G. J., Algra, A., de Kort, G. A., Witkamp, T. D., et al. (2015). DIAGRAM Investigators. Diagnostic yield and accuracy of CT angiography, MR angiography, and digital subtraction angiography for detection of macrovascular causes of intracerebral haemorrhage: prospective, multicentre cohort study. *BMJ* 351:h5762. doi: 10.1136/bmj.h5762

Wagle, W. A., Smith, T. W., and Weiner, M. (1984). Intracerebral hemorrhage caused by cerebral amyloid angiopathy: radiographic-pathologic correlation. *AJNR Am. J. Neuroradiol.* 5, 171–176.

Wu, P. S., Li, H., Zeng, N. Y., and Li, F. P. (2022). FMD-Yolo: an efficient face mask detection method for COVID-prevention and control in public. *Image Vis. Comput.* 117:104341. doi: 10.1016/j.imavis.2021.104341

Xie, H., Ma, S., Wang, X., and Zhang, X. (2020). Noncontrast computer tomography-based radiomics model for predicting intracerebral hemorrhage expansion: preliminary findings and comparison with conventional radiological model. *Eur. Radiol.* 30, 87–98. doi: 10.1007/s00330-019-06378-3

Xu, W., Ding, Z., Shan, Y., Chen, W., Feng, Z., Pang, P., et al. (2020). A Nomogram Model of Radiomics and Satellite Sign Number as Imaging Predictor for Intracranial Hematoma Expansion. *Front. Neurosci.* 14:491. doi: 10.3389/fnins.2020.00491

Zeng, N. Y., Li, H., and Peng, Y. H. (2021). A new deep belief network-based multi task learning for diagnosis of Alzheimer's disease. *Neural. Comput. Appl.* 1–12. doi: 10.1007/s00521-021-06149-6

Zeng, N. Y., Qiu, H., Wang, Z. D., Liu, W. B., Zhang, H., and Li, Y. R. (2018). A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing* 320, 195–202. doi: 10.1016/j.neucom.2018.09.001

Zeng, N. Y., Wu, P. S., Wang, Z. D., Li, H., Liu, W. B., and Liu, X. H. (2022). A small sized object dection oriented multi-scale feature fusion approach with application to defect detection. *IEEE Trans. Instrum. Meas.* 71, 1–14.

Zhang, Y., Zhang, B., Liang, F., Liang, S., Zhang, Y., Yan, P., et al. (2019). Radiomics features on non-contrast-enhanced CT scan can precisely classify VMH-related hematomas from other spontaneous intraparenchymal hematoma types. *Eur. Radiol.* 29, 2157–2165. doi: 10.1007/s00330-018-5747-x

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

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

Copyright © 2022 Wang, Xiong, Ye, Yang, He, Liu and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Neuroscience

TYPE Review
PUBLISHED 03 October 2022
DOI 10.3389/fnins.2022.878750

Check for updates

OPEN ACCESS

EDITED BY
Yuanpeng Zhang,
Nantong University, China

REVIEWED BY
Seyed Kamaledin Setarehdan,
University of Tehran, Iran
Felix Scholkmann,
University Hospital of Zürich, Switzerland
Paola Pinti,
Birkbeck, University of London, United Kingdom
Noman Naseer,
Air University, Pakistan

*CORRESPONDENCE
Keum-Shik Hong
kshong@pusan.ac.kr

SPECIALTY SECTION
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

RECEIVED 18 February 2022
ACCEPTED 06 July 2022
PUBLISHED 03 October 2022

CITATION
Huang R, Hong K-S, Yang D and Huang G (2022) Motion artifacts removal and evaluation techniques for functional near-infrared spectroscopy signals: A review. *Front. Neurosci.* 16:878750. doi: 10.3389/fnins.2022.878750

COPYRIGHT
© 2022 Huang, Hong, Yang and Huang. 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.

Motion artifacts removal and evaluation techniques for functional near-infrared spectroscopy signals: A review

Ruisen Huang[1], Keum-Shik Hong[1,2]*, Dalin Yang[3] and Guanghao Huang[4]

[1]School of Mechanical Engineering, Pusan National University, Busan, South Korea, [2]Department of Cogno-Mechatronics Engineering, Pusan National University, Busan, South Korea, [3]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO, United States, [4]Institute for Future, School of Automation, Qingdao University, Qingdao, China

With the emergence of an increasing number of functional near-infrared spectroscopy (fNIRS) devices, the significant deterioration in measurement caused by motion artifacts has become an essential research topic for fNIRS applications. However, a high requirement for mathematics and programming limits the number of related researches. Therefore, here we provide the first comprehensive review for motion artifact removal in fNIRS aiming to (i) summarize the latest achievements, (ii) present the significant solutions and evaluation metrics from the perspective of application and reproduction, and (iii) predict future topics in the field. The present review synthesizes information from fifty-one journal articles (screened according to three criteria). Three hardware-based solutions and nine algorithmic solutions are summarized, and their application requirements (compatible signal types, the availability for online applications, and limitations) and extensions are discussed. Five metrics for noise suppression and two metrics for signal distortion were synthesized to evaluate the motion artifact removal methods. Moreover, we highlight three deficiencies in the existing research: (i) The balance between the use of auxiliary hardware and that of an algorithmic solution is not clarified; (ii) few studies mention the filtering delay of the solutions, and (iii) the robustness and stability of the solution under extreme application conditions are not discussed.

KEYWORDS

functional near-infrared spectroscopy, motion artifacts removal, hemodynamic response, filtering techniques, noise suppression, signal-to-noise ratio

Introduction

Functional near-infrared spectroscopy (fNIRS) is a non-invasive brain imaging technique that uses near-infrared light (typically of wavelengths between 650 and 1,000 nm) to monitor hemodynamics changes in the cortical layer. Compared to electroencephalography (EEG), fNIRS enables to measure brain-activity related hemodynamics in terms of cerebral oxygenation and is less susceptible to electric noises (Huppert et al., 2009; Tak and Ye, 2014; Naseer and Hong, 2015; Chiarelli et al., 2017; Afkhami et al., 2019; Ghafoor et al., 2019; Khan et al., 2021). In addition, fNIRS can be integrated into a portable, wearable, and ergonomic device at low costs and operational expenses, making it a superior candidate for a user-friendly brain-computer interface system compared to other modalities, such as functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) (Hu et al., 2010; Piper et al., 2014; Scholkmann et al., 2014a; Pinti et al., 2015; Wyser et al., 2017; Zhao and Cooper, 2018; Hong and Zafar, 2018; Zhao H. B. et al., 2020, 2021; Ghafoor et al., 2021; Huang and Hong, 2021).

Fantini et al. (1999) reported that the artifacts caused by subjects' movements, specifically motion artifacts (MAs), can significantly influence the quality of the recorded optical signals of fNIRS. Some studies claimed that MAs reduce the signal-to-noise ratio (SNR) of fNIRS signals (Izzetoglu et al., 2005; Izzetoglu et al., 2010). Researchers have also verified that using MA removal techniques can ameliorate classification accuracy in cognition experiments (Zhou et al., 2021). Therefore, the issue concerning the causes, characteristics, and rejection methods of MAs in fNIRS signals is a crucial topic in fNIRS studies (Safaie et al., 2013; Piper et al., 2014).

During the early days, researchers skipped the analysis or discarded the data set when the measured signals were significantly corrupted by motion artifacts (Bartocci et al., 2000; Jasdzewski et al., 2003; Akgul et al., 2005; Khan and Hong, 2015; Nguyen and Hong, 2016; Zafar and Hong, 2017). Schroeter et al. (2003) removed the outliers manually. Moosmann et al. (2003) eliminated the disturbances by immobilizing the subjects' heads with a vacuum pad. Subsequently, a moving average was used (Kameyama et al., 2004; Lee et al., 2007). Channel rejection is another common method in early studies (Wilcox et al., 2005; Blasi et al., 2010). Attempts were made to remove MAs using an improved optical model. Nevertheless, the performances of the aforementioned solutions were not sufficient (Scholkmann et al., 2014b). Nowadays, a correction for motion artifacts has become a common consensus: Some processed the signals in two stages: artifact identifications and artifact corrections in the existing methods (Scholkmann et al., 2010; Virtanen et al., 2011; Yucel et al., 2014a), or minimized a user-defined cost function (Kim et al., 2011), or proposed a new model to compensate the artifacts (Izzetoglu et al., 2010; Yamada et al., 2015).

The existing literature indicates that the movements that cause MAs are diverse. Several studies have reported that the movements of subjects' heads (including nodding, shaking, tilting, etc.) could result in MAs in fNIRS measurements (Izzetoglu et al., 2005; Radhakrishnan et al., 2009; Robertson et al., 2010; Kim et al., 2011; Cui et al., 2015). Some researchers further discovered that facial muscle movement, including raising eyebrows, can lead to MAs (Izzetoglu et al., 2005; Robertson et al., 2010; Yucel et al., 2014b; Zhou et al., 2021). In addition, body movements, including the movement of the upper and lower limbs, degrade the fNIRS signals by causing head movements or by the inertia of the device (Yucel et al., 2014a; Rea et al., 2014; Abtahi et al., 2017; Vitorio et al., 2017; Khan et al., 2018; Siddiquee et al., 2020; Dybvik and Steinert, 2021). Vinette et al. (2015) monitored five epilepsy patients for a long period. Their data showed that MAs existed when the subjects were talking, eating, or drinking. These behaviors involve jaw movements. Novi et al. (2020) found that jaw movements could lead to two different motion artifacts. The direct cause of MAs is imperfect contact between the optodes and the scalp, including displacement, non-orthogonal contact, and oscillation of the optodes (Yamada et al., 2015; Nishiyori, 2016).

In this study, the authors reviewed journal articles concerning MA removal techniques in the Web of Science database. The keywords and numbers of journal papers are listed in **Table 1**. To narrow down our review scope, we list all journal articles found in the database. Subsequently, all the overlapping papers and irrelevant articles were removed by examining their context, which yielded 89 papers. Next, we proceeded to select journal papers that satisfied at least one of the following criteria: (i) The paper proposes a novel MA removal technique; (ii) the paper presents a quantitative comparison of several MA removal techniques; and (iii) the paper introduces a toolbox for MA removal. Eventually, 55 papers were selected from the literature. Forty-three papers presented a new solution to suppress MAs, seven papers compared the performance of the existing methods, and one study introduced a toolbox. **Figure 1** shows the partitioning of different types of papers in the selection process. Among the 47 new solutions, twelve solutions added auxiliary hardware. A list of selected studies and their categories is presented in **Table 2**. Since research on MA

TABLE 1 The number of journal papers (1990∼2022) obtained from the Web of Science database by combining different keywords.

Keywords	Paper number
fNIRS + motion artifact	90
NIRS + motion artifact	61
NIRS + motion correction	18
fNIRS + motion correction	27
Functional near-infrared spectroscopy + motion correction	32
Near-infrared spectroscopy + motion correction	51
Near-infrared spectroscopy + motion artifact	150
Functional near-infrared spectroscopy + motion artifact	109

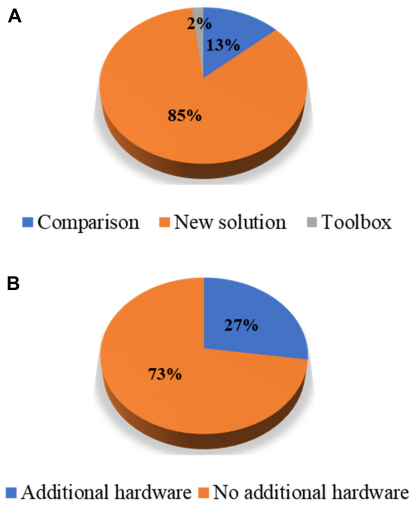


FIGURE 1
Percentage partitions: **(A)** Article types of the selected papers and **(B)** hardware-based solutions against algorithmic solutions among the papers proposing new solutions.

removal techniques requires a solid foundation in mathematics and programming, it is difficult for new scholars to assimilate the existing solutions in their studies. Moreover, some solutions were described in the text rather than using equations, making it difficult for other researchers to reproduce the reported methodologies. Therefore, this study aims to (i) provide a general view of the latest achievements in MA removal studies, (ii) briefly introduce several significant solutions from the view point of application and reproduction by using equations, and (iii) discuss future topics in the field.

This study is divided into five sections. The "Introduction" section presents the causes and significance of MA issues in fNIRS. In addition, this section specifies the objectives of this study and provides a quantitative summary of the existing literature on the topics. The section "Additional hardware-based techniques" summarizes the existing hardware-based solutions. The section "Signal processing-based techniques" discusses the algorithmic solutions. The section "Evaluation metrics" briefly introduces the definitions of some metrics to evaluate the performance of MA removal techniques. The final section "Conclusions and outlook" concludes this study and discusses potential issues concerning MA removal. We will use the compact notation provided in **Table 3** for the remainder of the paper.

Additional hardware-based techniques

Among the eight-nine selected papers, 17 discussed solutions using additional hardware, while 11 studies discussed accelerometer-related methods. Other auxiliary hardware includes a headpost cemented to the skull, a three-dimensional (3D) motion capture system, collodion-fixed prism-based optical fibers, an inertia measurement unit (IMU), a gyroscope, a magnetometer, and a camera. This section presents two solutions using accelerometers and one using linearly polarized light.

Accelerometer

Accelerometer-based methods include adaptive filtering, active noise cancelation (ANC) (Kim et al., 2011), accelerometer-based motion artifact removal (ABAMAR) (Virtanen et al., 2011), acceleration-based movement artifact reduction algorithm (ABMARA) (Metz et al., 2015), multi-stage cascaded adaptive filtering (Islam et al., 2017), blind source separation, accelerometer-based artifact rejection, and detection (BLISSA2RD) (von Luhmann et al., 2019). The introduction of the accelerometer improves the feasibility of real-time rejection of MAs.

Active noise cancelation

The method assumes that the measured signals, $z(n)$, are the sum of motionless signals, $x(n)$, and MAs, $v(n)$ (Kim et al., 2011). The objective of the solution is to minimize the power of the recovered signals, that is,

$$\begin{aligned}\min\left(\mathrm{E}\left(\hat{x}\left(n\right)^2\right)\right) &= \min\left(\mathrm{E}\left(\left(x\left(n\right)+v\left(n\right)-\hat{v}\left(n\right)\right)^2\right)\right)\\ &= \min\left(\mathrm{E}\left(x\left(n\right)^2\right)+2\mathrm{E}\left(x\left(n\right)v\left(n\right)\right)\right.\\ &\quad\left.-2\mathrm{E}\left(x\left(n\right)\hat{v}\left(n\right)\right)+\mathrm{E}\left(\left(v\left(n\right)-\hat{v}\left(n\right)\right)^2\right)\right).\end{aligned} \tag{1}$$

where $E(\cdot)$ denotes the expectation function, and the hat over x and v denotes the estimation of motionless fNIRS signals and MAs. Ideally, x is uncorrelated to either v or the estimate of v. Therefore, the two cross-terms on the right-hand side are equal to zero implying that the objective is equivalent to minimizing the square difference between the MAs and the estimated MAs. Moreover, because the actual MAs are highly correlated to the accelerometer output, $a(n)$, but unknown to users, $v(n)$ is replaced by $a(n)$ in the application. Subsequently, the final objective of signal processing is to minimize the square difference between $a(n)$ and the estimated MAs, that is,

$$\min\left(\mathrm{E}\left(\left(a\left(n\right)-\hat{v}\left(n\right)\right)^2\right)\right) \tag{2}$$

The estimate of $v(n)$ can be obtained using the difference between $z(n)$ and the estimate of $x(n)$. The estimate of $v(n)$ can then be computed in real time using a recursive least-squares filter. The procedure for the solution is graphically presented in **Figure 2**. The ANC solution was applied to optical intensities in real time. Whether this method was applied to optical densities or concentration changes is not clear. Another

TABLE 2 List of selected papers, article types, and information on additional hardware.

Paper	Type	Additional hardware
Izzetoglu et al. (2005)	New solution	
Zhang et al. (2005)	New solution	
Lee et al. (2007)	New solution	
Blasi et al. (2010)	New solution	Accelerometer
Cui et al. (2010)	New solution	
Izzetoglu et al. (2010)	New solution	
Robertson et al. (2010)	Comparison	
Scholkmann et al. (2010)	New solution	
Kim et al. (2011)	New solution	Accelerometer
Virtanen et al. (2011)	New solution	Accelerometer
Cooper et al. (2012)	Comparison	
Molavi and Dumont (2012)	New solution	
Amian and Setarehdan (2013)	New solution	
Barker et al. (2013)	New solution	
Santosa et al. (2013)	New solution	
Sweeney et al. (2013)	New solution	
Takahashi et al. (2013)	New solution	3D motion capture system
Brigadoi et al. (2014)	Comparison	
Selesnick et al. (2014)	New solution	
Yucel et al. (2014b)	New solution	Collodion-fixed prism-based optical fibers
Yucel et al. (2014a)	New solution	
Chiarelli et al. (2015)	New solution	
Hu et al. (2015)	Comparison	
Metz et al. (2015)	New solution	Accelerometer
Yamada et al. (2015)	New solution	Linearly polarized light sources, an orthogonally polarized analyzer
Barker et al. (2016)	New solution	
Gu et al. (2016)	New solution	
Guerrero-Mosquera et al. (2016)	New solution	
Islam et al. (2017)	New solution	Accelerometer
Janani and Sasikala (2017)	Comparison	
Santosa et al. (2017)	New solution	
Behrendt et al. (2018)	Comparison	
Dong and Jeong (2018)	New solution	
Jahani et al. (2018)	New solution	
Lee et al. (2018)	New solution	
Nguyen et al. (2018)	New solution	
Reyes et al. (2018)	Comparison	
Shukla et al. (2018)	New solution	
Siddiquee et al. (2018)	New solution	Inertia measurement unit (IMU)/accelerometer, gyroscope, magnetometer
Sutoko et al. (2018)	New solution	
Di Lorenzo et al. (2019)	Comparison	
Fishburn et al. (2019)	New solution	
Seghouane and Ferrari (2019)	New solution	
von Luhmann et al. (2019)	New solution	Accelerometer
Arunkumar and Bhaskar (2020)	New solution	
Gao et al. (2020)	New solution	
Sherafati et al. (2020)	New solution	
Zhao J. et al. (2020)	New solution	
Gemignani and Gervain (2021)	Comparison	
Lacerenza et al. (2021)	New solution	TD-fNIRS
Perpetuini et al. (2021)	New solution	Camera
Zhao Y. et al. (2021)	Toolbox	
Zhou et al. (2021)	New solution	
Gao L. et al. (2022)	New solution	
Kim et al. (2022)	New solution	
Huang et al. (2022)	New solution	
Gao Y. et al. (2022)	New solution	

issue in the approach of Kim et al. (2011) is that the performance was visually evaluated. Therefore, a quantitative evaluation of ANC in the sense of both noise suppression and signal distortion is needed.

Accelerometer-based motion artifact removal algorithm

The ABAMAR method is an offline analysis method for MA removal, where accelerometer outputs are used for MA detection, and the MA removal process is based on the measured fNIRS signals (Virtanen et al., 2011). Accordingly, we first define two Boolean functions as follows.

$$f_1(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3}$$

$$f_2(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{4}$$

A motion event can then be identified using the acceleration at the x- and y-axes, where Δt is the sampling period. The subscript x or y at $a(n)$ denotes the acceleration at the two axes. The subscript ME denotes the motion event, and operator $\vee$ denotes the OR operation. The flag for the motion events can be computed as follows.

$$\begin{aligned} flag_{\mathrm{ME}} = {} & f_1\left(|a_x(n) - a_x(n-1)| - 1.3\Delta T\right) \\ & \vee f_1\left(\left|a_y(n) - a_y(n-1)\right| - 1.3\Delta T\right) \end{aligned} \tag{5}$$

If $flag_{ME}$ is one, the signals encounter a motion event; otherwise, zero. Once a first true value of the flag appears, the timing of the motion event starts and is stored/defined as T_m. The motion event is ended when the flag remains false for over 20 s. The ending time is identified as the last sample when $flag_{ME}$ is true.

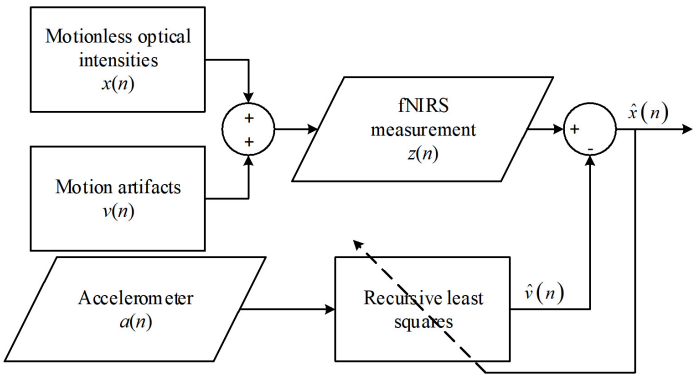


FIGURE 2
Procedures of active noise cancellation (ANC) algorithm.

Another flag, $flag_{MA}$, is introduced to identify the existence of MAs, which is defined as follows.

$$flag_{\mathrm{MA}} = f_1\left(T_m - 1\right) \tag{6}$$

T_m is the starting time of a motion event. The baseline of fNIRS signals, Avg, is defined as the average of the signal amplitudes before and after T_m. To avoid disturbance during the motion event, we marked 5 s before T_m as T_{before} and 5 s after T_m as T_{after}. The amplitudes of the signals before T_m, Avg_{before}, and those after T_m, Avg_{after}, are calculated as follows.

$$Avg_{\mathrm{before}} = \mathrm{mean}\left(z\left(n\right)|_{T_{\mathrm{before}}-15\le n<T_{\mathrm{before}}}\right) \tag{7}$$

$$Avg_{\mathrm{after}} = \mathrm{mean}\left(z\left(n\right)|_{T_{\mathrm{after}}<n\le T_{\mathrm{after}}+15}\right) \tag{8}$$

$flag_{BS}$ is a flag identifying baseline shifts in motion events. Specifically, one for baseline shifts, and zero for no shift.

$$flag_{\mathrm{BS}} = f_2\left(\left|Avg_{\mathrm{before}} - Avg_{\mathrm{after}}\right| - 2.6\sigma_{\mathrm{before}}\right) \tag{9}$$

$$\sigma_{\mathrm{before}} = \mathrm{std}\left(z\left(n\right)|_{T_{\mathrm{before}}-15\le n<T_{\mathrm{before}}}\right) \tag{10}$$

The function std(·) computes the standard deviation of its input.

The correction procedure only applies to baseline shift segments. Moreover, a flag for correction, $flag_{crr}$, is introduced. If its value is one, a correction of signals will be conducted; otherwise, it is zero. Moreover, operator $\wedge$ denotes the AND operation. The $flag_{crr}$ can be computed as follows.

$$\begin{aligned} flag_{\mathrm{crr}} = {} & f_1\left(N_{\mathrm{ch}}\left(flag_{\mathrm{BS}}|_{flag_{\mathrm{BS}}=1}\right) - 2\right) \\ & \vee f_1\left(N_{\mathrm{wv}}\left(flag_{\mathrm{BS}}|_{flag_{\mathrm{BS}}=1} - 2\right)\right) \wedge flag_{\mathrm{ME}} \end{aligned} \tag{11}$$

where $N_{ch}(\cdot)$ denotes the number of channels satisfying the condition specified in the input, and $N_{wv}(\cdot)$ denotes the number of wavelengths satisfying the input. When $flag_{crr}$ is one, $z(n)$ is corrected as follows.

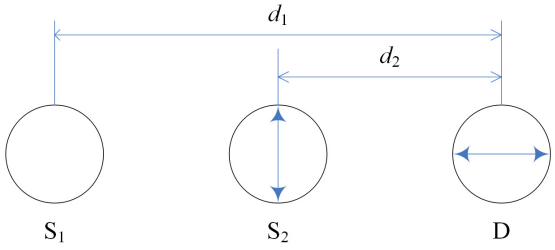


FIGURE 3
Optode arrangement for multidistance optode arrangement technique.

$$\hat{z}\left(n\right)\big|_{n\ \mathrm{after}\ T_m} = \frac{Avg_{\mathrm{before}}}{Avg_{\mathrm{after}}} z\left(n\right) \tag{12}$$

$$\hat{z}\left(n\right)\big|_{n\ \mathrm{inside}\ T_m} = A_{\mathrm{before}} \tag{13}$$

The ABAMAR solution is applied to optical intensities, optical densities, and concentration changes. It can efficiently suppress step-like artifacts. However, the signal details during motion events will be lost owing to the correction method. Moreover, empirical constants, such as 1.3 g/s in Eq. (5) (g denotes the gravitational acceleration of 9.81 m/s^2) and 2.6 in Eq. (9), may need to be updated for tasks other than sleeping monitoring. Some researchers have proposed an AMARA, which is an improvement of ABAMAR (Metz et al., 2015), by combining the movement artifact reduction algorithm (MARA; see section "Spline interpolation") and ABAMAR.

Linearly polarized light-based solution

Multidistance optode arrangement technique

An optical model of light transmission between a source and a detector is developed using light transmittance

TABLE 3 Definitions of variables, parameters, and their values.

Variables/ Parameters	Definitions
$z(n)$	Measured signals
$x(n)$	Motionless signals
$v(n)$	Motion artifacts
$\hat{x}(n)$	Estimated motionless functional near-infrared spectroscopy (fNIRS) signals
$\hat{v}(n)$	Estimated motion artifacts
$a(n)$	Accelerometer output
Δt	Sampling interval
$\vee$	OR operation
$flag_{ME}$	Flag for motion events
T_m	Timing of a motion event starts
T_{before}	5 s before T_m
T_{after}	5 s after T_m
Avg_{before}	Signals' amplitudes before T_m
Avg_{after}	Signals' amplitudes after T_m
$flag_{BS}$	Flag identifying the baseline shifts in motion events
std(·)	Standard deviation of its input
$flag_{crr}$	Flag for correction
$\wedge$	AND operation
$N_{ch}(\cdot)$	Number of channels satisfying a given condition
$N_{wv}(\cdot)$	Number of wavelengths satisfying a given input
G	Gravity acceleration approximating 9.81 m/s^2
$r_s(n)$	Light transmittances of the source-scalp gap
$r_d(n)$	Light transmittances of the detector-scalp gap
N	Time instance
I_0	Light intensity emitted by the source
$I_1(n)$	Light intensity reflected by hair
$I_2(n)$	Light intensity scattered by head tissue
DPF	Differential pathlength factor
D	Source-detector distance
$\Delta\mu$	Absorption coefficients change in the gray matter
$g(n)$	Wiener filter
$p_x(w)$	Power spectral densities (PSD) of the actual fNIRS signals
$p_v(w)$	PSD of the motion artifacts
a_i	Parameters of the AR model
$\phi(n)$	Composed of p motion artifact-free fNIRS signals
ω_n	Zero-mean noises in the AR model
Q	Error covariance matrix
ν_n	Measurement noise with an error covariance matrix R
W	Samples of moving time window
$msd(n)$	Moving standard deviation
$\xi(n)$	Corresponding samples
$z_{MA}(n)$	Motion artifacts segments
$z_{NC}(n)$	Non-corrupted segments
$z_{diff}(n)$	Difference between $z_{MA}(n)$ and its spline interpolation fitting
i_0	Coarsest scale
$\Phi(n)$	Mother scaling function
$\Psi(n)$	Mother wavelet function
N	Number of samples in the signals
NCDF(·)	Normal cumulative distribution function
σ_{HbO}	Standard deviation of z_{HbO}
σ_{HbR}	Standard deviation of z_{HbR}
W_L	Large-window size
W_S	Small-window size
N_{ch}	Number of channels
Id	An identity matrix
ν_1 *and* ν_2	Two types of motion artifacts
Ω	Gaussian white noise
$[\cdot]_{i,j}$	Element located at the ith row and the jth column
$[\cdot]_i$	ith element of the vector
W_1	Window size of the first median filter
W_2	Window size of the second median filter
Var (·)	Variance of the variable in the parentheses
σ_{dur}	Standard deviation of the filtered fNIRS signals during stimulus
σ_{pre}	Standard deviation of the filtered fNIRS signals before stimulus
$\in$	Small nonnegative constant
A_m	Set of time segments where motion artifacts (Mas) occur
N_{trl}	Number of trials in the signals
N_{sub}	Number of subjects
ΔCC	Difference of correlation coefficient

(Yamada et al., 2015). Here, the light transmittances of the source-scalp gap, detector-scalp gap, and head tissue at time instance n are denoted by $r_s(n)$, $r_d(n)$, and $R(n)$, respectively. Moreover, the light intensity emitted by the source is denoted by I_0, the light intensity reflected by the hair is denoted by $I_1(n)$, and the light intensity scattered by the head tissue is denoted by $I_2(n)$. The optical density can then be computed as follows.

$$\begin{aligned}\Delta A\,(n) &= -\frac{I_1\,(n)+I_2\,(n)}{I_1\,(0)+I_2\,(0)}\\ &= -\log\frac{I_1\,(n)+I_0 r_s\,(n)\,R\,(0)\,r_d\,(n)}{I_1\,(0)+I_0 r_s\,(0)\,R\,(0)\,r_d\,(0)}\\ &\quad -\log\frac{I_1\,(n)+I_0 r_s\,(n)\,R\,(n)\,r_d\,(n)}{I_1\,(n)+I_0 r_s\,(n)\,R\,(0)\,r_d\,(n)},\end{aligned}\tag{14}$$

where

$$I_0 = I_1\,(n)+I_2\,(n)\tag{15}$$

The MA removal solution includes two steps: Step 1 involves suppressing $I_1(n)$, and Step 2 involves attenuating the first term in the model.

Polarized optical films are attached to the source and detector in an orthogonal direction to suppress $I_1(n)$. Light reflection does not change the polarization of light, but scattering will change; therefore, only scattered light can be captured by the detector, that is, $I_1(n) = 0$. Thus, Eq. (14) can be reduced to the following form:

$$\Delta A(n) = -\log \frac{r_s(n)\, r_d(n)}{r_s(0)\, r_d(0)} - \log \frac{R(n)}{R(0)} \tag{16}$$

Step 2 is to cancel out the first term because it is independent of the hemodynamic changes. The optode arrangement is depicted in **Figure 3**. Accordingly, the method assumes that if two unidirectional inline channels have a small distance difference (two sources and one detector), their concentration changes will have similar temporal patterns. The hemodynamic changes contribute to the second term in Eq. (16), so according to the modified Beer–Lambert law, we obtain the following equation:

$$-\log \frac{R(n)}{R(0)} = DPF \cdot d\Delta\mu \tag{17}$$

where DPF denotes the differential pathlength factor, d indicates the source-detector distance, and $\Delta\mu$ corresponds to the absorption coefficient change in the gray matter. Moreover, the light transmittances of the source-scalp gaps for both channels and the detector-scalp gap for the detector are denoted by $r_{s1}(n)$, $r_{s2}(n)$, and $r_d(n)$, respectively. $R_1(n)$ and $R_2(n)$ denote the transmittances of the head tissues for the two channels. We can obtain the following equation from Eq. (17) by weighted subtraction of the optical densities of the two channels.

$$\begin{aligned}\Delta A_1(n) - k\Delta A_2(n) &= -\log \frac{r_{s1}(n)\, r_d(n)}{(r_{s2}(n)\, r_d(n))^k} \\ &\quad - \log r_{s1}(0)\, r_d(0) + k \log r_{s2}(0)\, r_d(0) \\ &\quad - \log \frac{R_1(n)}{R_1(0)} + k \log \frac{R_2(n)}{R_2(0)} \\ &= -\log \frac{r_{s1}(n)\, r_d(n)}{(r_{s2}(n)\, r_d(n))^k} \\ &\quad + C + \left(DPF_1 d_1 - kDPF_2 d_2\right) \Delta\mu(n).\end{aligned} \tag{18}$$

The constant k (approximately one) depends on the wavelength, and C is a constant that depends on the initial installation of the device. When the two sources and the detector are fixed relatively well, $r_{s1}(n)$ and $r_{s2}(n)$ will be consolidated to a similar value. Thus, the first term tends to be zero.

The solution above was evaluated on a hairy phantom, and its hardware solution in the first step inspired a creative solution to the hair-blocking problem encountered while using fNIRS devices. The validity of the second step is decided by the approximation that $k = 1$, whereas in actual measurement, k may occasionally become negative (Yamada et al., 2015). Besides, the optical model neglects detection noise and the angular fluctuation of optodes. The multidistance optode arrangement technique is applied to optical density and is available for real-time monitoring. The solution in Step 1 (using polarized optical film) can attenuate hair-reflected light.

Signal Processing-Based Techniques

Wiener filter

The Wiener filter approach is the first to remove motion artifacts without incorporating additional hardware devices (Izzetoglu et al., 2005; Orihuela-Espina et al., 2010; Li L. et al., 2021). The technique assumes that the measured fNIRS signals are a simple addition between the actual fNIRS signals, $x(n)$, and motion artifacts, $v(n)$. Moreover, it is assumed that $x(n)$ and $v(n)$ are stationary and uncorrelated.

$$\operatorname{corr}(x(n), v(n)) = 0 \tag{19}$$

Consequently, the Wiener filter, $g(n)$, minimizes the mean square error between $x(n)$ and $\hat{x}(n)$, that is,

$$\min\left(\mathrm{E}\left[e(n)^2\right]\right) = \min\left(\mathrm{E}\left[\left(x(n) - \hat{x}(n)\right)^2\right]\right) \tag{20}$$

Therefore, the optimum filter can be obtained using the orthogonality principle and simplified using Eq. (19) as follows.

$$\begin{aligned}\operatorname{corr}[e(n), x(n) + v(n)] &= \operatorname{corr}[x(n), x(n)] \\ &\quad - g(n) * \operatorname{corr}[x(n) + v(n), \\ &\quad x(n) + v(n)] \\ &= 0.\end{aligned} \tag{21}$$

We obtain the Fourier transform of the Wiener filter by converting Eq. (21) into the frequency domain using the Fourier transform, which is as follows.

$$G(w) = \frac{p_x(w)}{p_x(w) + p_v(w)} \tag{22}$$

where $p_x(w)$ and $p_v(w)$ denote the power spectral densities (PSDs) of the actual fNIRS signals and motion artifacts.

In application, a prior experiment is required to determine the PSDs of $g(n)$ by calculating the values of $x(n)$ and $v(n)$. Subsequently, we can determine $g(n)$ on a time scale using the inverse Fourier transform and apply it to new experimental data.

The Wiener filter is the first attempt to remove motion artifacts without a reference signal from additional hardware devices, such as accelerometers. With the $g(n)$ determined, the filter can be implemented for online applications. However, it requires prior knowledge of the PSDs of $x(n)$ and $v(n)$, which makes initial calibration more complex (a particularly designed paradigm is needed). The idea of building the filter model from a prior experiment inspired later research on motion artifact removal techniques.

Kalman filter

The Kalman filter approach was also proposed based on the general idea of the Wiener filter but as a different model (Izzetoglu et al., 2010). The motion artifact-free fNIRS signal, $x(n)$, was modeled using an autoregressive (AR) model. An AR model of order p can be written as

$$x(n) = \sum_{i=1}^{p} a_i x(n-i) \tag{23}$$

With the motion artifact-free data from the prior experiment, the parameters of the AR model, a_i, $i = 1, \ldots, p$, can be determined using the Yule-Walker equations. Therefore, the process equation for the Kalman filter has the following form.

$$\phi(n) = A\phi(n-1) + \omega_n,\ \phi(n) = \left[x(n) \cdots x(n-p+1) \right]^T \tag{24}$$

where $\phi(n)$ is composed of p motion artifact-free fNIRS signals, and ω_n denotes the zero-mean noises in the AR model with an error covariance matrix Q. Matrix A can be obtained using Eq. (23) as follows.

$$A = \begin{bmatrix} a_1 & \cdots & a_{p-1} & a_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \tag{25}$$

The measurement equation can be written as follows.

$$z(n) = C\phi(n) + v_n \tag{26}$$

$$C = \underbrace{\left[1\ 0 \cdots 0 \right]}_{p \text{ elements}} \tag{27}$$

where v_n denotes the measurement noise (such as the motion artifact) with an error covariance matrix R. $z(n)$ denotes the motion artifact corrupted signal.

Subsequently, Eqs. (24) and (26) form the state-space model for the Kalman filter, see **Table 4**. The minus sign in the superscript denotes the prior estimate of a variable. Subsequently, the motion-free fNIRS signal can be obtained using the Kalman filter (Wan and Nelson, 2001). The Kalman filter method can be applied to any online application of optical intensities, optical densities, and concentration changes, and in the Kalman filter theory, both ω_n and v_n are assumed to be zero-mean Gaussian white noise (Huang and Hong, 2021; Haghighi and Pishkenari, 2021; Li B. A. et al., 2021; Sun and Zhao, 2021; Yang et al., 2021; Lv et al., 2021; Tang et al., 2020; Pham et al., 2021). However, it is not the case for v_n (motion artifacts do not observe the zero-mean Gaussian distribution). It may degrade the filter's performance (Zhou et al., 2017). Moreover, matrices A and C were fixed once determined by the Yule-Walker method. Therefore, further development of the algorithm focuses on; (i) the compensation of the instrumental noises (Amian and Setarehdan, 2013) and (ii) adaptive adjustment of A and C, or using a more sophisticated and nonlinear model in place of Eqs. (24) and (26) (Dong and Jeong, 2018).

Additionally, the state-space model for the Kalman filter method is not unique. A typical example is the incorporation of the autoregressive iterative robust least-squares model (AR-IRLS), the general linear model (GLM), and two linear Kalman filters (Barker et al., 2016). A GLM and an AR-IRLS replaced the state-space model in the Kalman filters. The GLM was introduced to describe the dynamics of hemodynamic responses and physiological noises. The AR-IRLS was introduced as compensation for the MAs in the signals. A difference between the GLM-based method and the AR model-based method is that the GLM-based method requires the information of experimental paradigms, while the AR model-based method does not. Despite the significant adaptation ability of the Kalman filter regarding its state-space model, its applications are limited due to the effort required to set its initial parameters (e.g., the error covariance matrices for the state and the observation).

TABLE 4 Kalman filter algorithm.

Initialization:

$\hat{\phi}(0) = E(\phi(0))$,

$P(0) = \mathrm{E}\left((\phi(0) - \mathrm{E}(\phi(0)))(\phi(0) - \mathrm{E}(\phi(0)))^T\right)$.

Computation: For $n = 1, 2, \ldots$, compute:

State estimate propagation

$\hat{\phi}^-(n) = A\hat{\phi}(n-1)$.

Error covariance propagation

$P^-(n) = AP(n-1)A^T + Q$.

Kalman gain matrix

$G(n) = P^-(n)C^T\left(CP^-(n)C^T + R\right)^{-1}$.

State estimate update

$\hat{\phi}(n) = \hat{\phi}^-(n) + G(n)\left(z(n) - C\hat{\phi}^-(n)\right)$.

Error covariance update

$P(n) = \left(Id - G(n)C\right)P^-(n)$.

Spline interpolation

The spline interpolation method was first proposed by Scholkmann's group and is referred to as the MARA (Scholkmann et al., 2010; Selb et al., 2015; Lee et al., 2021). The method made two fundamental assumptions: (i) The measured fNIRS signal is a linear addition of the motion artifacts and the motion-free fNIRS signal, and (ii) in the motion-corrupted segments in the signal, the motion artifact component dominates the measured fNIRS signal. Therefore, the proposed MARA comprised two parts: (i) Motion artifact detection and

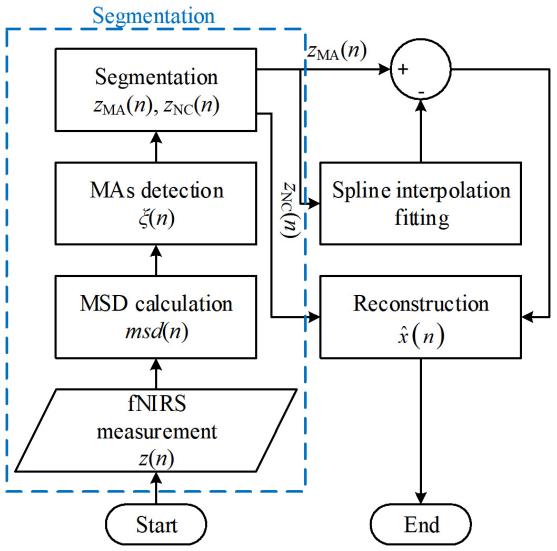


FIGURE 4
Flowchart of the movement artifact reduction algorithm (MARA) algorithm. The process blocks in the blue box are one of the reasons that limit the solution's online application.

segmentation, and (ii) motion artifact removal. A flowchart of MARA is illustrated in **Figure 4**. The spline interpolation method encompasses six processing steps. First, the moving standard deviation (MSD) is calculated within a moving time window of W samples and stored as $msd(n)$.

$$msd(n) = \frac{1}{W}\sqrt{\sum_{i=-k}^{k} n+i - \frac{1}{W}\left(\sum_{i=-k}^{k} z(n+i)\right)^2}, \quad W = 2k+1,\ k \in \mathbb{N}^* \tag{28}$$

where N^* is the set of natural numbers.

We can determine the start and end points of the motion artifacts and store the indices of the corresponding samples in a vector $\xi(n)$ by comparing with the MSD [i.e., $msd(n)$ in Eq. (28)] with a user-defined threshold value T. If the MSD is smaller than T, the corresponding $msd(n)$ will be assigned as zero. The start points and endpoints of motion artifacts can be extracted by considering the first and last samples of non-zero values in $msd(n)$. Next, let us suppose that there are L segments of motion artifacts and let the motion artifact segments, $z_{MA}(n)$, and the non-corrupted segments, $z_{NC}(n)$, be expressed, respectively, as

$$z_{\mathrm{MA}}(n) = \left\{z_{\mathrm{MA},1}(n), \cdots, z_{\mathrm{MA},L}(n)\right\} \tag{29}$$

$$z_{\mathrm{NC}}(n) = \left\{z_{\mathrm{NC},1}(n), \cdots, z_{\mathrm{NC},L}(n)\right\} \tag{30}$$

Using Eqs. (29) and (30), the measured fNIRS signals can be segmented into non-corrupted segments and motion artifact segments.

In the next part, the spline interpolation method corrects motion artifact segments. Because the motion artifact components dominate the MA segments, the spline interpolation fitting of $z_{MA}(n)$ can be viewed as the motion artifact component. The difference between $z_{MA}(n)$ and its spline interpolation fitting is stored as $z_{diff}(n)$:

$$z_{\mathrm{diff}}(n) = \left\{z_{\mathrm{diff},1}(n), \cdots, z_{\mathrm{diff},L}(n)\right\} \tag{31}$$

Because $z_{NC}(n)$ and $z_{diff}(n)$ may have different magnitude levels, the final step involves correcting the signal levels for the entire time series. Each segment is parallel-shifted according to the mean of the previous segment and that of the target segment. Two empirical constant thresholds, $\alpha = 3^{-1}\ \mathrm{Hz}^{-1}\cdot fs$ and $\beta = 2\ \mathrm{Hz}^{-1}\cdot fs$ (where the variable fs denotes the sampling frequency), were chosen for comparison. The detailed shifting rules are listed in Table 1 in Scholkmann et al. (2010).

The spline-interpolation method is used widely for offline analysis in fNIRS studies. It has also been included in some open-source toolboxes, such as HomER2 and NIRSLAB (Balardin et al., 2017). Moreover, the method applies not only to the optical intensities and optical densities but also to concentration changes. However, both the segmentation procedure (the procedures in the blue box in **Figure 4**) and the parallel-shifting procedure (in the reconstruction process) increase the difficulty for online filtering applications. Moreover, the filter performance depends on artifact detection results (Brigadoi et al., 2014). Some variations of the spline interpolation method have also been proposed (Jahani et al., 2018; Zhou et al., 2021).

Wavelet-based method

The wavelet-based method eradicates motion artifacts by removing the corresponding wavelet coefficients (Molavi and Dumont, 2012; Pinti et al., 2015), without requiring auxiliary devices. The method made the same assumption as the Wiener filter case, that is, $z(n) = x(n) + v(n)$. Based on discrete wavelet transform (DWT), fNIRS signals can be expanded as follows.

$$z(n) = \sum_{k} a_{i_0 k}\varphi_{i_0 k}(n) + \sum_{i=i_0}^{\infty}\sum_{k} b_{ik}\psi(n) \tag{32}$$

where i denotes the dilation parameter, k indicates the translation parameter, and i_0 denotes the coarsest scale. The scaling function $\varphi_{i_0 k}$ and the wavelet function ψ_{ik} are as follows.

$$\varphi_{i_0 k}(n) = 2^{i_0/2}\Phi\left(2^{i_0}n-k\right) \tag{33}$$

$$\psi_{ik}(n) = 2^{i/2}\Psi\left(2^{i}n-k\right) \tag{34}$$

The functions $\Phi(n)$ and $\Psi(n)$ correspond to the mother scaling function and mother wavelet function, respectively. With the fast wavelet transform, the coefficients of the DWT of $z(n)$ become

$$a_{i_0k} = \sum_{l=0}^{N-1} h\left(l-2k\right) z\left(l\right), k = 0, \ldots, 2^i - 1 \quad (35)$$

$$b_{ik} = \sum_{l=0}^{N-1} g\left(l-2k\right) z\left(l\right), i = i_0, \ldots, J-1 \quad (36)$$

where N denotes the number of samples in the signal and $N = 2^J$. $g(n)$ and $h(n)$ denote the wavelet filter bank high- and low-pass filters, respectively. Therefore, Eqs. (35) and (36) can be written in the matrix form as follows.

$$W = \Omega Z, Z = \left[z\left(0\right) \ldots z\left(N-1\right) \right]^T \quad (37)$$

where $W = [B_1, B_2, \ldots, B_{J-1}, A_{i0}]^T$, $B_i = [b_{i0}, \ldots, b_{i(2\wedge i-1)}]$, $A_{i0} = [a_{i00}, \ldots, a_{i0(2\wedge i0-1)}]$. Ω is an N by N DWT matrix. For any element w_{ik} in the wavelet coefficient vector W, the probability of observing its value can be written as

$$p_{ik} = 2\left(1 - NCDF\left(\frac{|w_{ik}|}{\hat{\sigma}}\right)\right) \quad (38)$$

where $NCDF(\cdot)$ is a normal cumulative distribution function. The estimated standard deviation of w_{ik} can be approximated empirically as

$$\hat{\sigma}_i = \frac{\text{Median}\left(|W_i|\right)}{0.6745} \quad (39)$$

where W_i denotes B_i or A_{i0} according to the value of i. Given a probability threshold α, if $p_{ik} < \alpha$, we consider that the corresponding wavelet coefficient w_{ik} is dominated by artifacts and set to zero. The filtered signals can be obtained from the updated wavelet coefficient vector as

$$\left[\hat{x}\left(0\right) \ldots \hat{x}\left(N-1\right) \right]^T = \Omega^T \hat{W} \quad (40)$$

The hat over $x(\cdot)$ denotes the filtered fNIRS signal, and the hat over W denotes the updated wavelet coefficient vector after thresholding.

Similar to the spline interpolation method, the wavelet-based method can be applied to optical intensities, optical densities, and concentration changes. Initially, the wavelet-based method was first introduced as an offline artifact removal algorithm. It may be adjusted to be an online method using the windowing technique; however, its filtering performance will be degraded. Several variations of this method have also appeared in recent years (Chiarelli et al., 2015; Shukla et al., 2018; Perpetuini et al., 2021). Gao Y. et al. (2022) addressed that the tuning of the probability threshold is crucial, and Wei et al. (2018) designed a dual-threshold structure to improve the performance of the wavelet-based method.

Correlation-based signal improvement method

Based on the negative correlation between measured oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) concentrations, the correlation-based signal improvement (CBSI) method was proposed in 2010 (Cui et al., 2010; Haeussinger et al., 2014). The algorithm made three fundamental assumptions: (i) HbO concentrations are strictly negatively correlated with HbR concentrations (close to –1); (ii) motion artifacts have identical effects on HbO and HbR, subject to a constant factor; and (iii) motion artifacts are not correlated to the actual concentration changes. The measured concentration changes were modeled as follows.

$$\begin{aligned} z_{\text{HbO}}\left(n\right) &= x_{\text{HbO}}\left(n\right) + \chi v, \\ z_{\text{HbR}}\left(n\right) &= x_{\text{HbR}}\left(n\right) + v, \end{aligned} \quad (41)$$

where v denotes the motion artifact of the HbR signals, and χ is a constant. Based on the first assumption, we can reasonably obtain

$$x_{\text{HbO}}\left(n\right) = -\beta x_{\text{HbR}}\left(n\right) \quad (42)$$

β denotes the free ratio accounting for the magnitude difference between HbO and HbR. Equations (41) and (42) can be used to obtain the following:

$$\begin{aligned} v\left(n\right) &= \frac{z_{\text{HbO}}(n)+\beta z_{\text{HbR}}(n)}{\chi+\beta}, \\ x_{\text{HbO}}\left(n\right) &= \frac{\beta}{\chi+\beta}\left(z_{\text{HbO}} - \chi z_{\text{HbR}}\right). \end{aligned} \quad (43)$$

Based on the third assumption, we set the correlation between v and x_{HbO} to zero, yielding

$$\begin{aligned} &\sum_n z_{\text{HbO}}\left(n\right)^2 + \left(\beta - \chi\right) \sum_n z_{\text{HbO}}\left(n\right) z_{\text{HbR}}\left(n\right) \\ &\quad - \chi\beta \sum_n z_{\text{HbR}}\left(n\right)^2 = 0 \end{aligned} \quad (44)$$

We assumed that $\chi = \beta$. Then, Eq. (44) can be reduced to

$$\chi = \frac{\sigma_{\text{HbO}}}{\sigma_{\text{HbR}}} \quad (45)$$

where σ_{HbO} and σ_{HbR} denote the standard deviations of z_{HbO} and z_{HbR}, respectively. Therefore, the final CBSI filter is

$$\begin{aligned} x_{\text{HbO}}\left(n\right) &= \frac{1}{2}\left(z_{\text{HbO}}\left(n\right) - \chi z_{\text{HbR}}\left(n\right)\right), \\ x_{\text{HbR}}\left(n\right) &= -\frac{1}{\alpha} x_{\text{HbO}}\left(n\right). \end{aligned} \quad (46)$$

Once χ is decided based on prior-experimental data, the CBSI filter can be easily implemented for online applications using Eq. (46). As mentioned in (Cui et al., 2010), CBSI can be applied after an exponential moving average (EMA) filter to facilitate real-time analysis. The behavior of an EMA filter is similar to an online version of a band-pass filter. The EMA is described as

follows.

$$\begin{aligned} L(n) &= \tfrac{1}{W_L} z(n) + \left(1 - \tfrac{1}{W_L}\right) L(n-1), \\ S(n) &= \tfrac{1}{W_S} \textstyle\sum_{k=0}^{W_S} z(n-k), \\ x(n) &= S(n) - L(n). \end{aligned} \tag{47}$$

$L(n)$ denotes the signal filtered by a large-window moving average (window size as W_L), and $S(n)$ denotes the signal filtered by a moving average of window size W_S.

Owing to the fundamental assumptions of the filter, the CBSI filter can only be applied to concentration changes. However, the negative correlation assumption between the HbO and HbR concentrations may not always be true for experimental data. Typically, the CBSI method is used for comparison in the literature (Janani and Sasikala, 2017; Reyes et al., 2018; Sherafati et al., 2020).

Principal component analysis

The solution based on principal component analysis (PCA) was developed from a multivariate method to remove physiological noises. The PCA-based method, entitled targeted PCA (tPCA), is similar to spline interpolation, except for using PCA instead of spline interpolation to fit the noisy segments in the signals (Yucel et al., 2014a). In this subsection, we subsequently explain the mechanism of Zhang's method and tPCA.

In accordance with the regular PCA method, the systemic spatial interference subspace (the actual hemodynamic responses) dominates in the baseline over the measured concentration changes and it spans during the stimulation periods for the first few components. The measured fNIRS signals are first converted to HbO and HbR concentration changes and are band-pass filtered. The filtered data from a chosen channel set are formatted into two $N_{ch} \times N$ matrices, Z_{HbO} and Z_{HbR}, where N_{ch} denotes the number of channels and N is the number of time samples. When segmented into two parts according to the baseline and stimulation periods, the two matrices turn out to be as follows.

$$Z_{\mathrm{HbO}} = \left[Z_{\mathrm{HbO,base}} \; Z_{\mathrm{HbO,stim}} \right] \tag{48}$$

$$Z_{\mathrm{HbR}} = \left[Z_{\mathrm{HbR,base}} \; Z_{\mathrm{HbR,stim}} \right] \tag{49}$$

For simplification, we omit the subscripts HbO and HbR, primarily because the filtering operations are identical for both chromophores. Accordingly, the spatial correlation matrices for the baseline signals are as follows.

$$C_{\mathrm{base}} = 1/N \cdot Z_{\mathrm{base}} Z_{\mathrm{base}}^T \tag{50}$$

The eigen-values and the eigenvectors can be obtained by the eigen-decomposition of spatial correlation matrices:

$$C_{\mathrm{base}} = U_{\mathrm{base}} \Sigma_{\mathrm{base}} U_{\mathrm{base}}^T \tag{51}$$

where the rows of U_{base} contains the eigenvectors, and Σ_{base} contains the eigenvalues in its diagonal. Based on the dominant interference subspace assumption, we assume that the first r spatial eigenvectors span in the stimulation period as $U_{base,r}$, an N_{ch} by r matrix padding by zeros. Accordingly, the actual hemodynamic responses will be the projection of the measured fNIRS signals onto the orthogonal subspace of the first r eigenvectors. The superscript $^{\perp}$ denotes the orthogonal subspace. Then, the filtered fNIRS signals were

$$X_{\mathrm{stim}} = \left(Id - U_{\mathrm{base},\,r} U_{\mathrm{base},\,r}^T \right) Z_{\mathrm{stim}} = U_{\mathrm{base},r}^{\perp} U_{\mathrm{base},\,r}^{\perp\,T} Z_{\mathrm{stim}} \tag{52}$$

where X_{stim} denotes the filtered fNIRS signal and matrix Id is an N_{ch} by N_{ch} identity matrix. Finally, we can block the average X_{stim} around the stimulus to improve the signal-to-noise ratio (SNR).

To remove MAs, tPCA applies the multivariate PCA method on MA segments. The eigenvectors in U_{base} are ranked in decreasing order of the percent variance of eigenvalues in Σ_{base}. The number of spatial eigenvectors span, r, is the minimum integer obtaining not less than 97% percent variance. Then, the filtered fNIRS signals of the segments are stitched back into the original signals after shifting the average of adjacent data segments (identical to MARA). The processing procedures are repeated twice or three times until there is no further improvement in the results.

Targeted PCA applies to optical densities. Due to the iteration in MA identification, the method cannot apply to online filtering. tPCA can remove both spike-like MAs and baseline shifts. It can better suppress high-frequency MAs than MARA (Yucel et al., 2014a). However, during application, the performance of tPCA may degrade if the MAs are not present in multiple channels (Brigadoi et al., 2014). The method requires multiple channels, which is not an appropriate choice when the number of channels is limited.

Multi-channel regression

The idea of the method is to subtract a best-fit linear combination of the noise references from each signal (Robertson et al., 2010). For a k-channel regression, the measured signal is modeled as follows.

$$z = aH + x \tag{53}$$

where $z \in \mathbb{R}^{1 \times N}$, $a \in \mathbb{R}1 \times k$, $H \in \mathbb{R}^{(k+1) \times N}$, $x \in \mathbb{R}^{1 \times N}$. The row vector, a, is the weight for each reference channel in H. The regressor H is defined as follows.

$$H = \begin{bmatrix} 1 & \cdots & 1 \\ z_{r1}(1) & \cdots & z_{r1}(N) \\ \vdots & \vdots & \vdots \\ z_{rk}(1) & \cdots & z_{rk}(N) \end{bmatrix} \tag{54}$$

where z_{ri} ($i = 1, 2, \ldots, k$) is the measurement of the ith reference channel. It is recommended to take co-located channels as reference channels in priority. Therefore, a can be estimated using the least-squares method.

$$a = \left(H^T H\right)^{-1} H^T z \tag{55}$$

The motionless signal, x, can be obtained by subtracting aH from z.

The multi-channel regression method does not require an auxiliary device and applies to optical intensities, optical densities, and concentration changes. It was first designed as an offline filtering method, but it can apply to online applications simply by replacing the least-square estimation procedure with recursive least-square estimation. When different types of artifacts affect collocated channels differently, the multi-channel regression method may not remove MAs entirely.

Independent component analysis

Independent component analysis (ICA) can decompose a signal to the weighted sum of multiple independent sources statistically (Kohno et al., 2007; Santosa et al., 2013). Therefore, the ICA solutions belong to blind source separation methods (von Luhmann et al., 2019). The key to removing MAs using ICA is how to select the components relevant to motionless signals or MAs. Kohno et al. (2007) suggested identifying the components related to artifacts using the coefficient of spatial uniformity (CSU).

$$CSU_j = \left| \frac{\bar{s}_j}{\text{std}\left(s_j\right)} \right| \tag{56}$$

Additionally, Santosa et al. (2013) estimated motionless signals by weighted reconstructing from independent components according to the t-values between each component and the desired hemodynamic response. The desired hemodynamic response is the convolution between the canonical hemodynamic response and the experimental paradigm (a boxcar function).

The ICA applies to optical intensities, optical densities, and concentration changes. The input requirements may vary according to the reconstruction strategy. Some researchers concluded that the performance of ICA was not satisfactory in some applications because fNIRS signals contained both non-instantaneous and non-constant coupling, correlated noise, and source dependencies (von Luhmann et al., 2019).

Temporal derivative distribution repair

The temporal derivative distribution repair (TDDR) is an online filtering method based on the temporal derivative of fNIRS signals. Three basic assumptions were made for this method: Non-motion fluctuations are assumed to be normally distributed; most fluctuations are independent of motion artifacts; and the derivatives of motion artifacts dominate in the derivatives of fNIRS signals when they are present (Fishburn et al., 2019). Therefore, the algorithm can be divided into five steps.

Step 1: For time instance n, the variation in the fNIRS signal can be calculated as follows.

$$Dz\left(n\right) = z\left(n\right) - z\left(n-1\right) \tag{57}$$

where the D before $z(n)$ denotes the temporal derivative of the measured signal at instance n.

Step 2: The observation weight $w(n)$, corresponding to $Dz(n)$, is initialized as one.

Step 3: The observation weight is adaptively updated using a robust estimator, M-estimator, and Tukey's biweight function:

$$w\left(n\right) = \begin{cases} \left(1 - \sigma_s\left(n\right)^2\right)^2 & \sigma_s\left(n\right) < 1, \\ 0 & \text{otherwise}, \end{cases} \tag{58}$$

where $\sigma_s(n)$ denotes the scaled deviation at instance n. An estimate of $\sigma_s(n)$ can be obtained from the weighted mean of the fluctuations, the absolute residuals of the temporal derivative, and the median absolute residual scale for the normal distribution. Let μ denote the weight mean of fluctuations, $res(n)$ denote the absolute residuals of temporal derivatives, and $\sigma(n)$ represent the median absolute residuals scale. Subsequently, we can obtain the following:

$$\mu = \frac{1}{\sum_{i=1}^{n} w\left(i\right)} \sum_{i=1}^{n} w\left(i\right) Dz\left(i\right) \tag{59}$$

$$res\left(n\right) = \left| Dz\left(n\right) - \mu \right| \tag{60}$$

$$\sigma = 1.4826 \cdot \text{median}\left(res\right) \tag{61}$$

Repeat (54)–(57) and consider $\sigma_s(n) = \sigma(n)$ until μ converges.

Step 4: The filtered temporal derivative can be obtained using μ and $w(n)$ obtained in Step 3.

$$Dx\left(n\right) = w\left(n\right)\left(Dz\left(n\right) - \mu\right) \tag{62}$$

Step 5: The final filtered signal is the integration of corrected temporal derivatives.

$$x\left(n\right) = \sum_{i=1}^{n} Dx\left(n\right) \tag{63}$$

Importantly, the filter can be implemented offline and online. The only difference is that when implemented online, the sum of $w(n)$ and $w(n)Dz(n)$ should be stored in each loop to facilitate the calculation in Eq. (59). The TDDR algorithm can be used for concentration changes, optical intensities, and optical densities. The performance and computation time of the TDDR depend

on the convergence criterion in Step 3. In particular, in Eq. (61), the iterative application of the median operator may result in a significant computation time. Furthermore, as the total number of data increases, the computation time increases according to regarding the complexity of the sorting algorithm in the median filter. Fishburn's group claimed that 130 ms were required to process a dataset of 20 Hz (50 ms in period) (Fishburn et al., 2019). A long computation time may lead to a loss in data transfer or a delay in the data output. In addition, high-frequency noise may degrade the performance of TDDR.

Transient artifact reduction algorithm

The transient artifact reduction algorithm (TARA) is an offline artifact removal method. The algorithm attempts to estimate motion artifacts instead of filtering the original signal directly (Selesnick et al., 2014). Generally, motion artifacts can be divided into two types. The Type 1 artifacts are modeled as spikes, whereas Type 2 artifacts are modeled as additive step shapes. The corrupted fNIRS signals were modeled as a low-pass filter/compound noise denoising problem (LPF/CSD).

$$z = x + \nu_1 + \nu_2 + \omega \tag{64}$$

where ν_1 and ν_2 denote the two types of MAs, and ω denotes the Gaussian white noise. Besides, the desired uncorrupted fNIRS signal x is assumed to approximate an all-zero signal when filtered by an appropriately chosen high-pass filter, denoted by HF. Accordingly, when designed as a zero-phase recursive discrete-time filter, H can be factorized into matrices A and B such that

$$HF = BA^{-1} \tag{65}$$

The first-order difference operator matrix D is defined as follows.

$$D = \underbrace{\begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 \quad 1 \end{bmatrix}}_{N \text{ elements}} \tag{66}$$

Inversely, the discrete-time integrator S is defined as the matrix satisfying $DS = Id_{N \times N}$, where $Id_{N \times N}$ indicates an N by N identity matrix. Then, matrix B can be further factorized as

$$B = B_1 D \tag{67}$$

Subsequently, we can change the variables using the following equations.

$$\nu_1 = Au_1 \tag{68}$$

$$\nu_2 = SAu_2 \tag{69}$$

Type 2 MAs are expressed in Eq. (69) because their first-order difference resembles Type 1 artifacts.

The two types of artifacts can be estimated using the following procedures when the factorization matrices A and B for the desired high-pass filter, two weight parameters λ_1 and λ_2, and two penalty functions r_0, r_1, and r_2 are available.

Step 1: Input fNIRS signals z, three regularization parameters λ_0, λ_1, and λ_2, penalty functions r_1 and r_2, and factorization matrices A and B for the desired high-pass filter (they can be calculated from a given degree, cutoff frequency, and length of z).

Step 2: The first additive terms for the estimator for both types of artifacts are as follows.

$$z_1 = B^T BA^{-1} z \tag{70}$$

$$z_2 = B_1^T BA^{-1} z \tag{71}$$

The vectors u_1 and u_2 will be initialized as zero vectors.

Step 3: Repeat the following calculation until u_1 and u_2 converge.

$$[\Lambda_0]_{n,n} = \lambda_0 / \psi_0 ([Au_1]_n) \tag{72}$$

$$[\Lambda_1]_{n,n} = \lambda_1 / \psi_1 ([DAu_1]_n) \tag{73}$$

$$[\Lambda_2]_{n,n} = \lambda_2 / \psi_2 ([Au_2]_n) \tag{74}$$

$$Q_1 = 2B^T B + A^T \left(\Lambda_0 + D^T \Lambda_1 D \right) A \tag{75}$$

$$Q_2 = 2B_1^T B_1 + A^T \Lambda_2 A \tag{76}$$

$$g = Bu_1 - B_1 u_2 \tag{77}$$

$$u_1 = Q_1^{-1} \left(z_1 + B^T g \right) \tag{78}$$

$$u_2 = Q_2^{-1} \left(z_2 + B_1^T g \right) \tag{79}$$

where Λ_i, (i = 0, 1, or 2) is diagonal, $\psi_i = u_i / r_i'$, $[\cdot]_{i,j}$ denotes the element located at the ith row and jth column of the matrix, and $[\cdot]_i$ denotes the ith element of the vector. The superscript prime notation, $'$, is the derivative operator.

Step 4: The estimated artifacts can be obtained using Eqs. (68) and (69). Accordingly, the filtered fNIRS signals becomes the following.

$$x = z - \nu_1 - \nu_2 \tag{80}$$

The TARA method is applied to optical densities, optical intensities, and concentration changes. TARA exhibits a similar performance to the wavelet-based method but is better at removing Type 2 motion artifacts. TARA is limited to Type 1 and Type 2 artifacts but is not designed for oscillatory transients.

Dual-stage median filter

The dual-stage median filter (DSMF) method resembles an online method for motion artifact removal (Huang et al., 2022). Similar to the TARA, the design of a DSMF is based on Type 1 and Type 2 artifact categories.

The measured fNIRS signal, z, is first filtered using the first median filter with a window size of W_1.

$$x_1 = \text{median}_{W_1}(z) \tag{81}$$

Subsequently, the difference between z and x_1 is filtered using the second median filter with a window size of W_2.

$$x_2 = \text{median}_{W_2}(z - x_1) \tag{82}$$

The output of the second median filter conserves the variations in the uncorrupted signals; however, it is biased. Therefore, the final step involves compensation of the bias using the initial time point.

$$x = x_2 + z(0) \tag{83}$$

W_1 can be set as 8 s, while W_2 as 18 s.

The DSMF is applied to optical intensities, optical densities, and concentration changes. Unlike the TDDR method, although two median filters are used in the algorithm, they are called once within fixed window sizes for each estimate. Moreover, registers can store the sorted data sequence during real-time applications, thereby reducing computation time. Therefore, a long computational time problem can be solved relatively quickly. Similar to the TARA, the DSMF is not designed for oscillatory transients.

Convolution neural networks

The convolution neural network (CNN) was adopted for MA removal in 2022 (Gao Y. et al., 2022; Kim et al., 2022). The denoising auto-encoder (DAE) model adopted a serial structure incorporating max-pooling and up-sampling layers (Gao Y. et al., 2022). The model enabled an offline MA removal by minimizing the linear combination of mean square error (L_{mse}), variance (L_{var}), standard deviation (L_{std}), and amplitude loss (L_{amp}).

$$Loss = L_{\text{mse}} + \theta_1 \times L_{\text{var}} + \theta_2 \times L_{\text{std}} + \theta_3 \times L_{\text{amp}} \tag{84}$$

where θ_1, θ_2, and θ_3 were set as 1, 1, and 10. The model parameters were estimated using simulated data generated from experimental data in the training process. The CNN structure is shown in **Figure 1** of the reference (Gao Y. et al., 2022).

Kim et al. (2022) adopted the CNN in the U-net structure for offline MA removal. Different from the DAE model, the U-net structure accepted an input containing two feature channels.

$$u(k) = \left[z(k)\ dHR(k) \right] \in \mathbb{R}^{N\times 2},\ k \in \{1, 2, \ldots, J\} \tag{85}$$

where $u(k)$ is the input of U-net of the kth package, J is the total number of packages, and $dHR(k)$ is the desired hemodynamic response of the kth package. $dHR(k)$ is the convolution between a canonical hemodynamic response function and a stimulus function.

$$dHR(k) = cHR * stim(k) \tag{86}$$

The U-net solution adopted the mean square errors between the ground-truth motionless fNIRS signals and the estimated signals as its loss function in the training process. The configurations of the U-net structure are shown in **Figure 1** of the reference (Kim et al., 2022).

The DAE model can be applied to optical intensities for offline filtering. It takes measured data in a full-time sequence, outputs the estimated motionless signals, and is trained using simulated data from an AR model. Therefore, the quality of the simulated data has a significant impact on the filtering performance of the model.

In contrast, the U-net structured CNN was applied to concentration changes. The training dataset was obtained through a semi-simulation process, and the size of each package was fixed as 1,024 samples. If the fNIRS data are obtained at a low sampling frequency, the package size is sufficient, but the filtering performance for high-sampling-frequency fNIRS data remains unknown. The matrix structure of the input data enables an easy adaptation to fNIRS measurement with multi-distance configurations. However, similar to the GLM-based method, the information of the stimulation paradigm is a prerequisite in the input. The inclusion of the stimulation paradigm is critical to obtain good performance of the method.

Comparison of diverse solutions

The methods presented above are the fundamental techniques available from the existing MA removal solutions. Among the new techniques under investigation, many solutions were developed from or in stack of one or several basic techniques. **Table 5** demonstrates the relationship between other new solutions and these basic techniques. For a fair comparison among these techniques, systematic approaches are required and the results are presented.

Comparisons have been made among the trial rejection method, recursive least squares adaptive filtering method (Park et al., 2020), wavelet-based method, ICA, two- or multi-channel regression method, Kalman filter, MARA, PCA, moving average, band-pass filter, median filter, Savitzky–Golay filter, and CBSI for adults' fNIRS data. Overall, the MA correction approach is better than the trial rejection approach (Brigadoi et al., 2014). MA correction can also help to improve classification accuracy (Janani and Sasikala, 2017). Wavelet-based method, ICA, multiple-channel regression, and MARA had a good performance in noise suppression. When processing children's

TABLE 5 List of motion artifact (MA) removal approaches and their relationship.

Paper	Solution name or category	Validation data type	Age group	Motion artifact type	Input	Source code availability
Izzetoglu et al. (2005)	Wiener filtering	Experimental	Adults		Optical intensities	Matlab
Zhang et al. (2005)	principal component analysis (PCA	Experimental	Adults		Concentration changes	Homer3
Lee et al. (2007)	Periodic moving average	Experimental	Adults		PPG signals	No
Blasi et al. (2010)	Channel rejection, adaptive filtering	Experimental	Infants		Concentration changes	Matlab
Cui et al. (2010)	Correlation-based signal improvement (CBSI)	Experimental	Adults	Spike	Concentration changes	Homer3
Izzetoglu et al. (2010)	Discrete Kalman filter	Experimental	Adults		Optical intensities	Matlab
Scholkmann et al. (2010)	Moving standard deviation, movement artifact reduction algorithm (MARA)	Experimental	Adults	Spike + drift	Optical densities	Homer3
Kim et al. (2011)	Active noise cancellation (ANC)	Experimental	Adults	Spike	Optical intensities	No
Virtanen et al. (2011)	Accelerometer-based motion artifact removal (ABAMAR)	Experimental	Adults	Spike + drift	Optical intensities and accelerations	No
Molavi and Dumont (2012)	Wavelet-based method	Experiment	Infants	Spike	Optical densities	Homer3
Amian and Setarehdan (2013)	ARMA model-based KF	Simulated + experimental	Adults		Optical intensities	No
Barker et al. (2013)	AR model and reweighted least squares	Simulated + experimental	Infants	Spike + drift	Optical intensities	No
Santosa et al. (2013)	Independent component analysis (ICA)	Experimental	Adults		Concentration changes	MNE-Python
Sweeney et al. (2013)	Ensemble empirical mode decomposition with canonical correlation analysis (EEMD-CCA)	Simulated + experimental	Adults		Optical intensities	No
Takahashi et al. (2013)	Motion artifact reconstruction	Experimental	Adults		Concentration changes	No

(Continued)

10.3389/fnins.2022.878750

TABLE 5 (Continued)

Paper	Solution name or category	Validation data type	Age group	Motion artifact type	Input	Source code availability
Selesnick et al. (2014)	Transient artifact reduction algorithm (TARA)	Simulated + experimental	Adults	Spike + drift	Optical intensities	https://ieeexplore.ieee.org/abstract/document/6942269/media#media
Yucel et al. (2014b)	Collodion-fixed prism-based optical fibers	Experimental	Adults	Spike + drift	Optical intensities	No
Yucel et al. (2014a)	Targeted PCA	Experimental	Adults	Spike + drift + oscillation	Optical intensities	Homer3
Chiarelli et al. (2015)	Kurtosis-based wavelet filtering (kbWF)	Semi-simulated	Adults	Spike	Optical densities	No
Metz et al. (2015)	MARA + ABAMAR	Experimental	Adolescents	Spike + drift	Optical densities + acceleration	No
Yamada et al. (2015)	Multidistance optode arrangement technique	Experimental	Phantom	Spike + drift	Optical intensities	No
Barker et al. (2016)	Kalman autoregressive, iterative robust least-squares functional near-infrared spectroscopy (fNIRS) model, KF	Semi-simulated	Adults	Spike	Optical intensities + stimulated function	AnalyzIR
Gu et al. (2016)	Empirical mode decomposition (EMD) + MARA	Semi-simulated	Children	Spike + drift	Optical intensities	No
Guerrero-Mosquera et al. (2016)	CBSI-based automatic artifact detection	Experimental	Adults		Concentration changes	No
Islam et al. (2017)	Multi-stage cascaded adaptive filtering (RLS) + singular spectrum analysis (SSA)	Experimental	Adults		Optical intensities + acceleration	No
Santosa et al. (2017)	Robust correlation of the innovations models	Simulated + experimental	Children		Optical intensities	www.bitbucket.org/huppertt/
Dong and Jeong (2018)	EKF with non-linear state-space model and short separation	Semi-simulated	Adults		Concentration changes	No
Jahani et al. (2018)	MARA, Savitzky-Golay filtering	Semi-simulated + experimental	Adults	Spike + drift	Optical densities	Homer3
Lee et al. (2018)	Wavelet-decomposed back-propagation neural network (BPNN) + contaminated channel identification algorithm based on entropy	Experimental	Adults	Spike + drift	Optical intensities	No
Nguyen et al. (2018)	Adaptive filtering based on RLS with an exponential forgetting factor	Simulated + experimental	Adults	Spike	Concentration changes	No
Shukla et al. (2018)	Stationary wavelet transforms, zero-mean Laplace distribution modeling	Experimental	Adults	Spike	Original EDA signals	No
Siddiquee et al. (2018)	Autoregression with exogenous input	Experimental	Adults		Concentration changes	No

(Continued)

10.3389/fnins.2022.878750

TABLE 5 (Continued)

Paper	Solution name or category	Validation data type	Age group	Motion artifact type	Input	Source code availability
Sutoko et al. (2018)	Adaptive algorithm, rejection	Simulated + experimental	Children		Product of concentration changes and optical path length	No
Fishburn et al. (2019)	TDDR + robust regression	Simulated + experimental	Children	Spike + drift + oscillation	Optical densities	MNE-Python
Seghouane and Ferrari (2019)	Robust fNIRS HRF estimation algorithm	Simulated + experimental	Adults		Optical intensities	No
von Luhmann et al. (2019)	Blind source separation and accelerometer-based artifact rejection and detection (BLISSA2RD), ICA + Canonical Correlation Analysis and temporal embedding	Simulated + experimental	Adults		Optical intensities	https://github.com/avolu/BLISSARD
Arunkumar and Bhaskar (2020)	Cascaded RLS, normalized least mean square (NLMS), LMS adaptive filter	Experimental	Adults		Optical intensities + acceleration	No
Sherafati et al. (2020)	Global variance of temporal derivatives (GVTD)	Experimental	Adults + infants		Optical densities or concentration changes	No
Zhao J. et al. (2020)	Targeted median filter and mathematical morphology (tMedMor)	Semi-simulated	Adults	Spike + drift + oscillation	Optical densities	No
Lacerenza et al. (2021)	TD-fNIRS	Experimental	Adults		Photon distribution of time-of-flight	No
Perpetuini et al. (2021)	Wavelet-based method, wavelet coherence (WCOH), video tracking	Semi-simulated	Adults	Spike + oscillation	Optical densities	No
Zhou et al. (2021)	NIRS-ICA toolbox, PCA	Simulated + experimental	Adults		Optical intensities	NIRS-ICA
Gao L. et al. (2022)	MARA + wavelet	Semi-simulated	Adults	Spike + drift + oscillation	Concentration changes	Homer3
Kim et al. (2022)	U-net CNN	Semi-simulated	Adults	Spike + drift + oscillation	Concentration changes	No
Huang et al. (2022)	Dual-stage median filter	Simulated + experimental	Adults	Spike + drift	Optical intensities	https://gitee.com/cognoholic/differential-median-filter.git
Gao Y. et al. (2022)	Deep learning-based CNN	Simulated + experimental	Adults	Spike + drift + oscillation	Concentration changes	https://github.com/YuanyuanGao216/fNIRS_denoise_by_DL

10.3389/fnins.2022.878750

fNIRS signals, researchers compared PCA, MARA, wavelet-based method, moving average, CBSI, and tPCA. MARA, tPCA, and CBSI can retain a higher number of trials. The tPCA and MARA are more robust. Moving average, wavelet-based method, and tPCA have a better performance than other methods. Researchers also examined the filtering performance of infants' fNIRS data. Trial rejection, wavelet-based method, tPCA, MARA, and different processing pipelines were compared in their studies. MA correction can retain many trials, but hemodynamic response functions were also suppressed. The optimal pre-processing pipeline depends on the dataset (Gemignani and Gervain, 2021). Similar to the conclusion for adults' data, MA correction is better than trial rejection. The wavelet-based method alone or in a stack with MARA and the wavelet-based method (iqr = 0.5) had better performance in noise suppression. MARA in a stack with the wavelet-based method can recover a higher number of trials.

Reportedly, the wavelet-based filter has good performance when MAs are caused by sudden large movements (or spikes). Moreover, ICA and multi-channel regression also perform well in terms of noise suppression (Robertson et al., 2010). Cooper et al. (2012) reported that PCA, MARA, wavelet-based method, Kalman filter, and trial rejection method could reduce the MSE in corrupted fNIRS signals (Cooper et al., 2012). PCA, MARA, and wavelet-based methods exhibited much better performance than the other methods (Piper et al., 2014). In contrast, Brigadoi et al. (2014) claimed that the Kalman filter performed better using the area under the curve (AUC ratio) of the mean hemodynamic responses than PCA, and MARA, followed by the wavelet-based method and CBSI. They also claimed that the wavelet-based method exhibited a better performance in reducing the AUC for the first two seconds after stimulus onset (AUC_{0-2}).

Therefore, the evaluation of MA removal techniques depends on the evaluation metrics of the researchers. Due to this phenomenon, multiple metrics need to be considered when comparing various methods. Cooper et al. (2012) used MSE to assess how different solutions can suppress MAs and contrast-to-noise ratio (CNR) to evaluate how well the solutions can retain brain activation signals. Janani and Sasikala (2017) reported that a band-pass filter, median filter, and Savitzky-Golay filter could increase the CNR. Additionally, the performance of the MA solutions differs from each other when the MA types are different. **Figure 5** showed an example of MA removal using different filters.

The optimal solution for different age groups can be different. Hu et al. (2015) tested the MA removal performance of six solutions using fNIRS data from children: wavelet-based method, PCA, MARA, moving average, CBSI, and the combination of moving average and wavelet-based methods. It was found that MARA successfully suppressed all types of MAs but had an unsatisfactory AUC ratio. Moving average and wavelet-based method turned out to be better solutions for children. Meanwhile, CBSI was efficient in removing spike-like and fast step-like MAs. Besides, Reyes et al. (2018) claimed that tPCA was the best solution for processing children's fNIRS signals and the authors added that CBSI occasionally produced unstable hemodynamic responses. However, the authors also claimed that MARA was robust and performed well in both AUC and the average standard deviation of each trial-specific hemodynamic response. The contradiction between Reyes and Hu's findings regarding the MARA originates from improper artifact correction (Hu et al., 2015). Another study by Di Lorenzo et al. (2019) used a combination of MARA and wavelet-based methods and found it to be more effective for infant data than using trial rejection, MARA, or wavelet alone. They further justified that using MA correction could retain many trials. Similar results were obtained by Behrendt et al. (2018). They claimed that trial rejection was not recommended for infants' data. **Table 6** summarizes the data type, age group, techniques in comparison, and main conclusions of all nine comparison papers.

Evaluation metrics

Various evaluation metrics have been defined in the literature for noise suppression and signal distortion to quantitatively evaluate the performance of artifact removal solutions. Some metrics have been defined to describe the extent to which the solutions could suppress the artifacts, while others were designed to assess the signal distortion after data processing. In this section, we introduce the details of several metrics for artifact removal except for some commonly used metrics such as root-mean-squared error (to assess the square-root distance between filtered signals and reference signals), the area under the curves (AUC. e.g., researchers can take the initial 1–2 s and the 2–6 s of HbO concentrations) (Reyes et al., 2018), and the *t*-test (e.g., used together with GLM) (Hu et al., 2015).

Noise suppression

Signal to noise ratio

The signal-to-noise ratio (ΔSNR) in motion artifact removal (Izzetoglu et al., 2005) is different from the more commonly known signal-to-noise ratio (SNR), which is simply the ratio between the power of real signals and the power of noises expressed in decibel. ΔSNR (ΔSNR) is defined as the difference between the estimation SNR (SNR_e) and the input SNR (SNR_i). The two variables are defined as follows.

$$SNR_e = 10\log_{10}\left(\frac{\mathrm{Var}\,(x)}{\mathrm{Var}\,(\hat{x}-x)}\right) \tag{87}$$

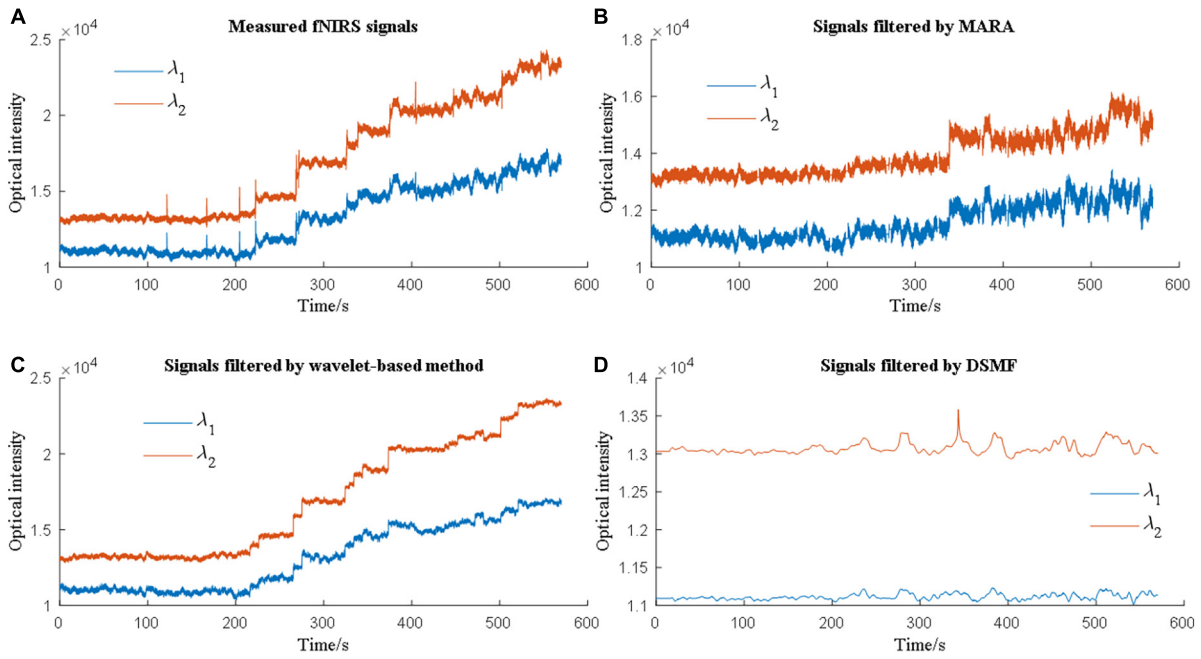


FIGURE 5
Different motion artifact (MA) removal techniques have different performances when processing experimental data.

$$SNR_i = 10\log_{10}\left(\frac{\mathrm{Var}\,(x)}{\mathrm{Var}\,(z-x)}\right) \tag{88}$$

The operator Var(·) denotes the variance of the variable in the parentheses. A hat over the motionless fNIRS signal x denotes the estimated motionless fNIRS signal, and z denotes the measured fNIRS signal. Subsequently, ΔSNR is defined as follows.

$$\Delta SNR = SNR_e - SNR_i \tag{89}$$

If the filter of interest can perfectly remove motion artifacts in z, then $\Delta SNR \to +\infty$. If $\Delta SNR < 0$, the filter of interest amplifies the artifacts in the measured signal. Otherwise, the filter can partially remove motion artifacts. In contrast, the larger the ΔSNR, the better the performance of the filter. Physically, the metric describes how well the filter can attenuate motion artifacts in the measured signals.

Since a motionless fNIRS signal is required in the calculation, ΔSNR can be well suited to evaluate the filtering performance of the simulated data. When applied to experimental data, a desired motionless signal is required in place of the motionless fNIRS signal.

Contrast to noise ratio

The contrast to noise ratio (CNR) describes the difference in the fNIRS signal before and during stimulation (Zhang et al., 2005). Here, the subscript "pre" denotes any signal before the stimulus, and the subscript "dur" denotes any signal during the stimulus. The CNR can then be computed as follows.

$$CNR = \frac{\bar{\hat{x}}_{\mathrm{dur}} - \bar{\hat{x}}_{\mathrm{pre}}}{\sqrt{\sigma^2_{\mathrm{dur}} + \sigma^2_{\mathrm{pre}}} + \epsilon} \tag{90}$$

where σ_{dur} and σ_{pre} denote the standard deviations of the filtered fNIRS signals during and before the stimulus, and ϵ corresponds to a small non-negative constant compensating for channels with a small stimulus response but relatively low activity. Empirically, $\epsilon = 0.1$ μM. In some studies, ϵ was ignored. When MAs exist, the CNR map may appear as all active or negative channels. The active and inactive areas spread throughout the map. Once MAs are properly removed, CNR maps with more concise active zones can be obtained (Kohno et al., 2007).

Percent root difference

The percent root difference (PRD) is a metric that describes how well the two signals match each other (Scholkmann et al., 2010; Dong and Jeong, 2018). When using the motionless signal x as a reference, the PRD is defined as follows.

$$PRD = 100\% \times \sqrt{\sum_{i=1}^{N} \frac{\left(\hat{x}\,(i) - x\,(i)\right)^2}{\sum_{i=1}^{N} x^2\,(i)}} \tag{91}$$

The smaller the PRD, the better the filtered signals match the motionless signals, that is, fewer artifacts exist in the filtered signals.

TABLE 6 Data type, age group, motion artifact (MA) removal techniques and main conclusions in the studies of MA comparison.

Paper	Data type	Age group	Techniques of interest	Main conclusions
Robertson et al. (2010)	Experimental	Adults	(i) RLS adaptive filtering (ii) Wavelet-based method (iii) ICA (iv) Two-channel regression (v) Multi-channel regression	(i) Independent component analysis (ICA) or multiple-channel regression has the largest SNR changes. (ii) RLS adaptive filtering produced the smallest SNR improvement. (iii) Wavelet-based method is most effective for spike-like artifacts. (iv) ICA and multiple-channel regression are good options when time-consuming manual techniques are absent.
Cooper et al. (2012)	Semi-simulated	Adults	(i) PCA (ii) Movement artifact reduction algorithm (MARA) (iii) Wavelet-based method (iv) Kalman filter	(i) All methods yield a significant reduction in mean square error (MSE) and an increase in contrast-to-noise ratio (CNR). (ii) MARA has the largest drop in MSE. (iii) Wavelet-based method has the highest average increase in CNR. (iv) The authors recommend the routine application of MA correction.
Brigadoi et al. (2014)	Experimental	Adults	(i) Trial rejection (ii) PCA (iii) MARA (iv) Wavelet-based method (v) Kalman filter (vi) Correlation-based signal improvement (CBSI)	(i) MA correction is better than trial rejection. (ii) Wavelet-based method is the most powerful method.
Hu et al. (2015)	Experimental	Children	(i) PCA (ii) MARA (iii) Wavelet-based method (iv) Moving average (v) CBSI (vi) Wavelet + moving average	(i) Moving average and wavelet-based method outstand the others. (ii) Types C and D artifacts have a larger influence on GLM model than Types A and B*.
Janani and Sasikala (2017)	Experimental	Adults	(i) Band-pass filter (ii) CBSI (iii) Median filter (iv) Savitzky–Golay filter (v) Wavelet denoising, ICA	(i) Wavelet-based method attenuates the MA energy and increases the CNR of Subjects 1 and 2. (ii) ICA also suppresses physiological noises and spike-like MAs. (iii) MA removal helps to classify motor tasks.
Behrendt et al. (2018)	Semi-simulated	Infants	(i) Trial rejection (ii) Wavelet-based method (iii) Targeted PCA (tPCA) (iv) tPCA + wavelet-based method	(i) The authors suggested using the wavelet-based method with iqr = 0.5. (ii) Trial rejection alone is not recommended. (iii) Assessing MA removal performance may not require full simulation studies.
Reyes et al. (2018)	Experimental	Children	(i) tPCA (ii) MARA (iii) Wavelet-based method (iv) CBSI	(i) tPCA, MARA, and CBSI retained a higher number of trials. (ii) The CBSI produced sometimes unstable hemodynamic response functions. (iii) The tPCA and MARA were the most robust in all metrics. (iv) The tPCA outperformed MARA. (v) The tPCA is an effective technique in MA correction for young children.
Di Lorenzo et al. (2019)	Semi-simulated + experimental	Infants	(i) Trial rejection (ii) MARA (iii) Wavelet-based method (iv) MARA + wavelet-based method	(i) MA correction is better than trial rejection. (ii) Wavelet-based method alone or in stack with MARA are the most effective in reducing the between- and within-subject standard deviation. (iii) MARA + wavelet-based method performs the best in semi-simulation and recovering the greatest number of trials.
Gemignani and Gervain (2021)	Simulated + experimental	Infants	Five pipelines	(i) MA correction can retain many trials. (ii) Hemodynamic response functions were suppressed as well. (iii) MA rejection preserved adequately the characteristics of HRF. (iv) The performance of all pipelines declined when noise power increased but less than no pre-processing. (v) No difference in running pre-processing on either optical densities or concentration changes. (vi) Pre-processing depends on the dataset.

*Type A: Spike-like MA with a standard deviation (SD) of 50 from the mean within 1 s. Type B: Spike-like MA with a SD of 100 from the mean within 1 to 5 s. Type C: Baseline shift with a gentle slope between 5 and 30 s with a SD of 300 from the mean. Type D: Slow baseline shift longer than 30 s with a SD of 500 from the mean.

10.3389/fnins.2022.878750

Artifact power attenuation

Artifact power attenuation (APA) indicates how much the artifacts in the measured signal can be suppressed in the signal of interest within the corrupted segments (Molavi and Dumont, 2012). Therefore, the metric is expressed in decibels as follows.

$$APA = 10\log_{10}\frac{\sum_{i\in A_m}\left(z^{\mathrm{HP}}(i)\right)^2}{\sum_{i\in A_m}\left(\hat{x}^{\mathrm{HP}}(i)\right)^2} \tag{92}$$

where A_m denotes the set of time segments in which the MAs occur. The set constrains the evaluation domain within corrupted segments. The superscript HP denotes a high-pass signal. It is introduced to remove low-frequency physiological variants (Molavi and Dumont, 2012). Since the measured signals are placed in the nominator, the larger the APA, the better the filter of interest that can suppress MAs. Typically, APA is expressed in decibels (Shukla et al., 2018).

APA can directly reveal a filter's resistance to MAs without disturbance from physiological noise. When using APA, the high-pass filter needs to be chosen carefully to ensure efficient filtering of the physiological noise. Importantly, the filters for both the nominator and denominator should be identical to facilitate a fair comparison.

Within/between-subject standard deviation

Within-subject standard deviation and between-subject standard deviation are two different metrics for evaluating the performance of the MA filter on the experimental data. Within-subject standard deviation refers to the average standard deviation of single-trial hemodynamic responses for a given channel and chromophore. N_{trl} denotes the number of trials in the signals, k denotes the sample index, i denotes the trial index, and N denotes the number of samples in a trial. $\Delta\hat{c}_i(k)$ is the estimated concentration change at instance k in trial i.

$$\sigma_{\mathrm{within}} = \frac{\sum_{k=1}^{N}\mathrm{std}_i\left(\Delta\hat{c}_i(k)\right)}{N},\ i = 1,\ldots,N_{\mathrm{trl}} \tag{93}$$

In practice, there is a within-subject standard deviation for each chromophore in each channel of each subject. The metric is used under the assumption that most of the variability between single-trial hemodynamic responses for each subject is because of the influence of MAs. When comparing the two signals, the within-subject standard deviations can be computed for each chromophore, channel, and subject. Then, the two values were mapped to a point in a scatter plot. We can compare the filtering performance of the two filters by counting the number of points below or above the line $y = x$.

The between-subject standard deviation refers to the standard deviation of the averaged hemodynamic responses across subjects for each channel (Brigadoi et al., 2014; Di Lorenzo et al., 2019). N_{sub} denotes the number of subjects, i denotes the subject number, N denotes the number of samples in an experiment, and $\Delta\hat{c}_i(k)$ denotes the estimated concentration change at instant k of subject i. The between-subject standard deviation becomes

$$\sigma_{\mathrm{between}} = \frac{\sum_{k=1}^{N_{\mathrm{sub}}}\mathrm{std}_i\left(\Delta\widehat{c}_i(k)\right)}{N},\ i = 1,\ \ldots,\ N_{\mathrm{sub}} \tag{94}$$

It is assumed that the MAs dominate the variability of hemodynamic responses across subjects when using the metric.

Signal distortion

Correlation coefficient

There are two methods for evaluating the motion artifact removal performance using the correlation coefficient. The first method computes the correlation between a reference signal and a block-average signal (Zhang et al., 2005). The reference signal was obtained by shifting the rectangular pulse signal to the right. The desired metric is then calculated as the correlation coefficient between the reference signal and the block-averaged response of the filtered signal.

The second method defines the difference in the correlation coefficient (ΔCC) (Izzetoglu et al., 2005; Jahani et al., 2018) as the difference between the estimated correlation coefficient and the input correlation coefficient, that is,

$$\Delta CC = \mathrm{corr}\left(x,\hat{x}\right) - \mathrm{corr}\left(x, z-x\right) \tag{95}$$

The operator corr(·, ·) represents the correlation between two variables in brackets. The second term in this definition tends to be zero. If $\Delta\mathrm{CC} < 0$, then the filter of interest significantly distorts the motionless signal. As ΔCC is closer to one, the distortion reduces in the filtered signal.

Similar to ΔSNR, ΔCC is well-suited for simulated data. When applied to experimental data, reference motionless signals are required in place of motionless fNIRS signals. In contrast, the first definition of the correlation metric requires an experimental paradigm, primarily because it compares the shifted rectangular pulse signal and the block-averaged response.

Pearson product-moment correlation coefficient

This metric appeared in 2008 and was designed to measure the similarity between two signals (Scholkmann et al., 2010). The coefficient is defined as follows.

$$r = \frac{1}{M}\sum_{i=1}^{N}\left(\frac{\hat{x}(i)-\bar{\hat{x}}(i)}{\sigma_{\hat{x}}}\right)\left(\frac{x(i)-\bar{x}(i)}{\sigma_x}\right) \tag{96}$$

where σ denotes the standard deviation and $M = N–1$. Moreover, the coefficient can also be used to evaluate the similarity between the motionless signal x and measured signal z. The only difference is the replacement of the estimated motionless signal with z in the equation. It is used to calculate the improvement of the proposed method compared with the measurement. The

TABLE 7 Suitability checking of the defined metrics to three types of data.

Metrics	Evaluation function	Simulated data	Semi-simulated data	Experimental data
ΔSNR	Noise suppression	√	√	
CNR	Noise suppression	√	√	√
PRD	Noise suppression	√	√	
APA	Noise suppression	√	√	√
Within-/between-subject SD	Noise suppression	√	√	√
ΔCC	Signal distortion	√	√	
Pearson product-moment correlation coefficient	Signal distortion	√	√	

more significant the absolute difference between $r(\hat{x}, x)$ and $r(x, z)$, the better the performance of the proposed method. As r becomes closer to one, there is less distortion in the signal of interest. Occasionally, the difference in the Pearson product-moment correlation coefficient can also be used to evaluate the improvement of the filtered signals. The difference is defined as follows.

$$\Delta r = r\left(\hat{x}, x\right) - r\left(z, x\right) \tag{97}$$

The difference may be expressed as a percentage.

Data type compatibility

According to Table 5, three types of data exist: (i) Simulated data, (ii) semi-simulated data, and (iii) experimental data. A semi-simulated dataset can be generated by adding desired concentration changes to MA-corrupted experimental data at resting state, or by adding uncorrupted experimental data at the resting state to the sum of desired concentration changes and designed artifact sequence. For both simulated and semi-simulated datasets, the ground-truth motionless signals are known. Nevertheless, this is not true for experimental datasets. Therefore, the input requirements of each metric decide the metric's compatible data types. Table 7 summarizes the data type compatibility of different metrics.

Conclusion and outlook

This review suggests that removing MAs is unavoidable in developing a more generalized fNIRS device. Literature reported that fNIRS demonstrate a good tolerance to motion artifacts when well-positioned (Pinti et al., 2020). However, a generalized fNIRS apparatus may not guarantee that all users will stick to this rule. The introduction of motion artifact removal techniques will add to the device's robustness and stability even when the users are involved in significant physical activities. These improvements will add to a positive user experience. The MA issue will become crucial for portable and wearable fNIRS devices. Once the problem is solved, we can envision a more exciting application of fNIRS to connect people or connect the real world with the digital world. Additionally, Tables 5 and 6 show that the filtering performance of one method on different age group can be divers. Hence, for a specific application, a study regarding the optimal solution is still necessary. Our in-depth review can help experienced researchers as well as novices in three ways. First, a brief introduction to some core solutions from a different perspective can help gain a better understanding as to how the solutions can be replicated, whether they are suitable for online or offline applications, and their advantages and disadvantages. Second, the definitions and physical meanings of commonly used metrics were summarized. Researchers can select the best metrics according to their requirements. Third, a sketch of the latest developments in MA removal techniques can help readers familiarize themselves with the field and decide their research focus. This review can be served as a list of different motion artifact removal solutions. When targeting the application of fNIRS, readers can pick the methods or metrics fulfilling their requirements for their specific scenario. When attacking a better noise removal solution, readers can develop their theory from the methods mentioned in this work or take another path.

The research in MA removal has made significant progress since the early 21st century. Different attempts have been made to remove MAs, especially for offline processing. Some researchers have also investigated the role of MA removal techniques in fNIRS data analysis pipelines (Gemignani and Gervain, 2021). Some studies managed to find out the optimal techniques among several candidates when doing experiments on a targeted population (Hu et al., 2015; Behrendt et al., 2018; Reyes et al., 2018; Di Lorenzo et al., 2019). Some attempted to solve the problem by improving the optical models (Scholkmann et al., 2014b; Yamada et al., 2015).

However, this review also revealed several lacunae concerning MA removal. First, no research has explained how to achieve a balance between the use of auxiliary hardware and that of an algorithmic solution. A comparative study

compared an adaptive filter (accelerometer is needed) with wavelet-based filtering, ICA, two-channel, and multi-channel regression (Robertson et al., 2010); however, it is far from sufficient. Second, the existing online MA removal solutions are primarily validated for noise suppression and signal distortion, but their filtering delays are not evaluated. An online solution with high noise suppression, low signal distortion, and low delay is significant for portable and wearable fNIRS devices. Third, the robustness and compatibility of the solutions in different application scenarios were not thoroughly studied. A solution's filtering performance may change when the user's age, tasks, external disturbance, or user's physical health are different. This issue is crucial for developing application-oriented fNIRS devices, such as devices that operate in extreme environments or a targeted population.

The research to remove MAs is not a method-oriented topic but an application-oriented topic. Any solution that can help make one step further to either general applications or a specific application would be advantageous to the fNIRS community.

Author Contributions

RH and K-SH designed the structure and scope of the review. RH aggregated reviewed articles and prepared the first draft manuscript. DY and GH verified the equations. K-SH reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the China Scholarship Council (grant no. 201808260086) and in part by the Technology Innovation Program (grant no. 20016846, Development of fNIRS-based multimonitoring healthcare device for sleep state diagnosis) funded by the Ministry of Trade, Industry & Energy, South Korea.

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 NN declared a past co-authorship with the author K-SH to the handling editor.

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

Abtahi, M., Amiri, A. M., Byrd, D., and Mankodiya, K. (2017). Hand motion detection in fNIRS neuroimaging data. *Healthcare* 5:5020020. doi: 10.3390/healthcare5020020

Afkhami, R. G., Walker, F. R., Ramadan, S., and Johnson, S. J. (2019). A dynamic model of brain hemodynamics in near-infrared spectroscopy. *IEEE Transac. Biomed. Eng.* 2019, 1–1. doi: 10.1109/TBME.2019.2954829

Akgul, C. B., Sankur, B., and Akin, A. (2005). Spectral analysis of event-related hemodynamic responses in functional near infrared spectroscopy. *J. Comput. Neurosci.* 18, 67–83. doi: 10.1007/s10827-005-5478-2

Amian, M., and Setarehdan, S. K. (2013). Motion artifact reduction in functional near infrared spectroscopy signals by autoregressive moving average modeling based Kalman filtering. *J. Innov. Opt. Health Sci.* 6:1350035. doi: 10.1142/S1793545813500351

Arunkumar, K. R., and Bhaskar, M. (2020). Heart rate estimation from wrist-type photoplethysmography signals during physical exercise. *Biomed. Sig. Proc. Cont.* 57:101790. doi: 10.1016/j.bspc.2019.101790

Balardin, J. B., Morais, G. A. Z., Furucho, R. A., Trambaiolli, L., Vanzella, P., Biazoli, C., et al. (2017). Imaging brain function with functional near-infrared spectroscopy in unconstrained environments. *Front. Hum. Neurosci.* 11:258. doi: 10.3389/fnhum.2017.00258

Barker, J. W., Aarabi, A., and Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomed. Optics Exp.* 4, 1366–1379. doi: 10.1364/BOE.4.001366

Barker, J. W., Rosso, A. L., Sparto, P. J., and Huppert, T. J. (2016). Correction of motion artifacts and serial correlations for real-time functional near-infrared spectroscopy. *Neurophotonics* 3:031410. doi: 10.1117/1.NPh.3.3.031410

Bartocci, M., Winberg, J., Ruggiero, C., Bergqvist, L. L., Serra, G., and Lagercrantz, H. (2000). Activation of olfactory cortex in newborn infants after odor stimulation: a functional near-infrared spectroscopy study. *Pediat. Res.* 48, 18–23. doi: 10.1203/00006450-200007000-00006

Behrendt, H. F., Firk, C., Nelson, C. A., and Perdue, K. L. (2018). Motion correction for infant functional near-infrared spectroscopy with an application to live interaction data. *Neurophotonics* 5, 015004. doi: 10.1117/1.NPh.5.1.015004

Blasi, A., Phillips, D., Lloyd-Fox, S., Koh, P. H., and Elwell, C. E. (2010). Automatic detection of motion artifacts in infant functional optical topography studies. *Oxy. Transport Tissue* 662, 279–284. doi: 10.1007/978-1-4419-1241-1_40

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., et al. (2014). Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *Neuroimage* 85, 181–191. doi: 10.1016/j.neuroimage.2013.04.082

Chiarelli, A. M., Maclin, E. L., Fabiani, M., and Gratton, G. (2015). A kurtosis-based wavelet algorithm for motion artifact correction of fNIRS data. *Neuroimage* 112, 128–137. doi: 10.1016/j.neuroimage.2015.02.057

Chiarelli, A. M., Zappasodi, F., Di Pompeo, F., and Merla, A. (2017). Simultaneous functional near-infrared spectroscopy and electroencephalography for monitoring of human brain activity and oxygenation: a review. *Neurophotonics* 4:041411. doi: 10.1117/1.NPh.4.4.041411

Cooper, R. J., Seib, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., et al. (2012). A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Front. Neurosci.* 6:147. doi: 10.3389/fnins.2012.00147

Cui, X., Baker, J. M., Liu, N., and Reiss, A. L. (2015). Sensitivity of fNIRS measurement to head motion: an applied use of smartphones

in the lab. *J. Neurosci. Methods* 245, 37–43. doi: 10.1016/j.jneumeth.2015.02.006

Cui, X., Bray, S., and Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49, 3039–3046. doi: 10.1016/j.neuroimage.2009.11.050

Di Lorenzo, R., Pirazzoli, L., Blasi, A., Bulgarelli, C., Hakuno, Y., Minagawa, Y., et al. (2019). Recommendations for motion correction of infant fNIRS data applicable to multiple data sets and acquisition systems. *Neuroimage* 200, 511–527. doi: 10.1016/j.neuroimage.2019.06.056

Dong, S., and Jeong, J. (2018). Improvement in recovery of hemodynamic responses by extended Kalman filter with non-linear state-space model and short separation measurement. *IEEE Transac. Biomed. Eng.* 66, 2152–2162. doi: 10.1109/TBME.2018.2884169

Dybvik, H., and Steinert, M. (2021). Real-world fNIRS brain activity measurements during Aashtanga Vinyasa yoga. *Brain Sci.* 11:742. doi: 10.3390/brainsci11060742

Fantini, S., Hueber, D., Franceschini, M. A., Gratton, E., Rosenfeld, W., Stubblefield, P. G., et al. (1999). Non-invasive optical monitoring of the newborn piglet brain using continuous-wave and frequency-domain spectroscopy. *Phys. Med. Biol.* 44, 1543–1563. doi: 10.1088/0031-9155/44/6/308

Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., and Medvedev, A. V. (2019). Temporal derivative distribution repair (TDDR): a motion correction method for fNIRS. *Neuroimage* 184, 171–179. doi: 10.1016/j.neuroimage.2018.09.025

Gao, L., Wei, Y. H., Wang, Y. F., Wang, G., Zhang, Q., Zhang, J. B., et al. (2022). Hybrid motion artifact detection and correction approach for functional near-infrared spectroscopy measurements. *J. Biomed. Optics* 27: 025003. doi: 10.1117/1.JBO.27.2.025003

Gao, Y., Cavuoto, L., Yan, P., Kruger, U., Schwaitzberg, S., De, S., et al. (2020). *A deep learning approach to remove motion artifacts in fNIRS data analysis. Biophotonics Congress: Biomedical Optics 2020 (Translational, Microscopy, OCT, OTS, BRAIN)*. Washington, DC: Optical Society of America, BM2C, 7.

Gao, Y., Chao, H., Cavuoto, L., Yan, P., Kruger, U., Norfleet, J., et al. (2022). Deep learning-based motion artifact removal in functional near-infrared spectroscopy. *Neurophotonics* 9:041406. doi: 10.1117/1.NPh.9.4.041406

Gemignani, J., and Gervain, J. (2021). Comparing different pre-processing routines for infant fNIRS data. *Dev. Cogn. Neurosci.* 48:100943. doi: 10.1016/j.dcn.2021.100943

Ghafoor, U., Lee, J. H., Hong, K.-S., Park, S. S., Kim, J., and Yoo, H. R. (2019). Effects of acupuncture therapy on MCI patients using functional near-infrared spectroscopy. *Front. Aging Neurosci.* 11:237. doi: 10.3389/fnagi.2019.00237

Ghafoor, U., Yang, D., and Hong, K.-S. (2021). Neuromodulatory effects of HD-tACS/tDCS on the prefrontal cortex: a resting-state fNIRS-EEG study. *IEEE J. Biomed. Health Inform.* 2021, 1–12. doi: 10.1109/JBHI.2021.3127080

Gu, Y., Han, J. X., Liang, Z. H., Yan, J. Q., Li, Z., and Li, X. L. (2016). Empirical mode decomposition-based motion artifact correction method for functional near-infrared spectroscopy. *J. Biomed. Optics* 21:15002. doi: 10.1117/1.JBO.21.1.015002

Guerrero-Mosquera, C., Borragan, G., and Peigneux, P. (2016). Automatic detection of noisy channels in fNIRS signal based on correlation analysis. *J. Neurosci. Methods* 271, 128–138. doi: 10.1016/j.jneumeth.2016.07.010

Haeussinger, F. B., Dresler, T., Heinzel, S., Schecklmann, M., Fallgatter, A. J., and Ehlis, A. C. (2014). Reconstructing functional near-infrared spectroscopy (fNIRS) signals impaired by extra-cranial confounds: an easy-to-use filter method. *Neuroimage* 95, 69–79. doi: 10.1016/j.neuroimage.2014.02.035

Haghighi, M. S., and Pishkenari, H. N. (2021). Real-time topography and hamaker constant estimation in atomic force microscopy based on adaptive fading extended kalman filter. *Internat. J. Cont. Automat. Syst.* 19, 2455–2467. doi: 10.1007/s12555-020-0076-7

Hong, K.-S., and Zafar, A. (2018). Existence of initial dip for BCI: an illusion or reality. *Front. Neurorob.* 12:69. doi: 10.3389/fnbot.2018.00069

Hu, X. S., Arredondo, M. M., Gomba, M., Confer, N., DaSilva, A. F., Johnson, T. D., et al. (2015). Comparison of motion correction techniques applied to functional near-infrared spectroscopy data from children. *J. Biomed. Optics* 20:126003. doi: 10.1117/1.Jbo.20.12.126003

Hu, X. S., Hong, K.-S., Ge, S. Z. S., and Jeong, M. Y. (2010). Kalman estimator- and general linear model-based on-line brain activation mapping by near-infrared spectroscopy. *Biomed. Eng. Online* 9:82. doi: 10.1186/1475-925X-9-82

Huang, R., and Hong, K.-S. (2021). Multi-channel-based differential pathlength factor estimation for continuous-wave fNIRS. *IEEE Access* 9, 37386–37396. doi: 10.1109/Access.2021.3063120

Huang, R., Qing, K., Yang, D., and Hong, K.-S. (2022). Real-time motion artifact removal using a dual-stage median filter. *Biomed. Sig. Proc. Cont.* 72:103301. doi: 10.1016/j.bspc.2021.103301

Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Optics* 48, D280–D298. doi: 10.1364/ao.48.00d280

Islam, M. T., Zabir, I., Ahamed, S. T., Yasar, M. T., Shahnaz, C., and Fattah, S. A. (2017). A time-frequency domain approach of heart rate estimation from photoplethysmographic (PPG) signal. *Biomed. Sig. Proc. Cont.* 36, 146–154. doi: 10.1016/j.bspc.2017.03.020

Izzetoglu, M., Chitrapu, P., Bunce, S., and Onaral, B. (2010). Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *Biomed. Eng. Online* 9:16. doi: 10.1186/1475-925x-9-16

Izzetoglu, M., Devaraj, A., Bunce, S., and Onaral, B. (2005). Motion artifact cancellation in NIR spectroscopy using Wiener filtering. *IEEE Transac. Biomed. Eng.* 52, 934–938. doi: 10.1109/Tbme.2005.845243

Jahani, S., Setarehdan, S. K., Boas, D. A., and Yucel, M. A. (2018). Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and Savitzky-Golay filtering. *Neurophotonics* 5:015003. doi: 10.1117/1.NPh.5.1.015003

Janani, A., and Sasikala, M. (2017). Investigation of different approaches for noise reduction in functional near-infrared spectroscopy signals for brain-computer interface applications. *Neur. Comp. Appl.* 28, 2889–2903. doi: 10.1007/s00521-017-2961-4

Jasdzewski, G., Strangman, G., Wagner, J., Kwong, K. K., Poldrack, R. A., and Boas, D. A. (2003). Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. *Neuroimage* 20, 479–488. doi: 10.1016/S1053-8119(03)00311-2

Kameyama, M., Fukuda, M., Uehara, T., and Mikuni, M. (2004). Sex and age dependencies of cerebral blood volume changes during cognitive activation: a multichannel near-infrared spectroscopy study. *Neuroimage* 22, 1715–1721. doi: 10.1016/j.neuroimage.2004.03.050

Khan, H., Naseer, N., Yazidi, A., Eide, P. K., Hassan, H. W., and Mirtaheri, P. (2021). Analysis of human gait using hybrid EEG-fNIRS-based BCI system: a review. *Front. Hum. Neurosci.* 14:613254. doi: 10.3389/fnhum.2020.613254

Khan, M. J., and Hong, K.-S. (2015). Passive BCI based on drowsiness detection: an fNIRS study. *Biomed. Optics Exp.* 6, 4063–4078. doi: 10.1364/BOE.6.004063

Khan, R. A., Naseer, N., Qureshi, N. K., Noori, F. M., Nazeer, H., and Khan, M. U. (2018). fNIRS-based neurorobotic interface for gait rehabilitation. *J. Neuroeng. Rehabilit.* 15:7. doi: 10.1186/s12984-018-0346-2

Kim, C.-K., Lee, S., Koh, D., and Kim, B.-M. (2011). Development of wireless NIRS system with dynamic removal of motion artifacts. *Biomed. Eng. Lett.* 1, 254–259. doi: 10.1007/s13534-011-0042-7

Kim, M., Lee, S., Dan, I., and Tak, S. (2022). A deep convolutional neural network for estimating hemodynamic response function with reduction of motion artifacts in fNIRS. *J. Neur. Eng.* 19:016017. doi: 10.1088/1741-2552/ac4bfc

Kohno, S., Miyai, I., Seiyama, A., Oda, I., Ishikawa, A., Tsuneishi, S., et al. (2007). Removal of the skin blood flow artifact in functional near-infrared spectroscopic imaging data through independent component analysis. *J. Biomed. Opt.* 12:062111. doi: 10.1117/1.2814249

Lacerenza, M., Spinelli, L., Buttafava, M., Dalla Mora, A., Zappa, F., Pifferi, A., et al. (2021). Monitoring the motor cortex hemodynamic response function in freely moving walking subjects: a time-domain fNIRS pilot study. *Neurophotonics* 8:015006. doi: 10.1117/1.NPh.8.1.015006

Lee, G., Jin, S. H., and An, J. (2018). Motion artifact correction of multi-measured functional near-infrared spectroscopy signals based on signal reconstruction using an artificial neural network. *Sensors* 18:2957. doi: 10.3390/s18092957

Lee, H. W., Lee, J. W., Jung, W. G., and Lee, G. K. (2007). The periodic moving average filter for removing motion artifacts from PPG signals. *Internat. J. Cont. Automat. Syst.* 5, 701–706.

Lee, S. H., Hur, S. W., Kwak, Y. Y., Nam, Y. H., and Kim, C. J. (2021). Ahead-time approach to carrot-chasing guidance law for an accurate trajectory-tracking control. *Internat. J. Cont. Automat. Syst.* 19, 2634–2651. doi: 10.1007/s12555-020-0413-x

Li, B. A., Tan, Y. H., Zhou, L., and Dong, R. L. (2021). Robust-nonsmooth Kalman filtering for stochastic sandwich systems with dead-zone. *Internat. J. Cont. Automat. Syst.* 19, 101–111. doi: 10.1007/s12555-019-1027-z

Li, L., Zhang, H., Nguyen, Q., and Hong, K.-S. (2021). Adaptive parameter identification for nonlinear sandwich systems with hysteresis nonlinearity based guaranteed performance. *Internat. J. Cont. Automat. Syst.* 19, 942–952. doi: 10.1007/s12555-019-2020-2

Lv, X., Fang, Y., Mao, Z., Jiang, B., and Qi, R. (2021). Fault detection for a class of closed-loop hypersonic vehicle system via hypothesis test method. *Internat. J. Cont. Automat. Syst.* 19, 350–362. doi: 10.1007/s12555-019-0906-7

Metz, A. J., Wolf, M., Achermann, P., and Scholkmann, F. (2015). A new approach for automatic removal of movement artifacts in near-infrared spectroscopy time series by means of acceleration data. *Algorithms* 8, 1052–1075. doi: 10.3390/a8041052

Molavi, B., and Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiolog. Meas.* 33, 259–270. doi: 10.1088/0967-3334/33/2/259

Moosmann, M., Ritter, P., Krastel, I., Brink, A., Thees, S., Blankenburg, F., et al. (2003). Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *Neuroimage* 20, 145–158. doi: 10.1016/s1053-8119(03)00344-6

Naseer, N., and Hong, K.-S. (2015). fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* 9:3. doi: 10.3389/fnhum.2015.00003

Nguyen, H. D., and Hong, K.-S. (2016). Bundled-optode implementation for 3D imaging in functional near-infrared spectroscopy. *Biomed. Opt. Exp.* 7, 3491–3507. doi: 10.1364/BOE.7.003491

Nguyen, H. D., Yoo, S. H., Bhutta, M. R., and Hong, K.-S. (2018). Adaptive filtering of physiological noises in fNIRS data. *Biomed. Eng. Online* 17:180. doi: 10.1186/s12938-018-0613-2

Nishiyori, R. (2016). fNIRS: an emergent method to document functional cortical activity during infant movements. *Front. Psychol.* 7:533. doi: 10.3389/fpsyg.2016.00533

Novi, S., Roberts, E., Spagnuolo, D., Spilsbury, B., Price, D. M., Imbalzano, C., et al. (2020). Functional near-infrared spectroscopy for speech protocols: characterization of motion artifacts and guidelines for improving data analysis. *Neurophotonics* 7:015001. doi: 10.1117/1.NPh.7.1.015001

Orihuela-Espina, F., Leff, D. R., James, D. R. C., Darzi, A. W., and Yang, G. Z. (2010). Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. *Phys. Med. Biol.* 55, 3701–3724. doi: 10.1088/0031-9155/55/13/009

Park, B., Pham, P., and Hong, K.-S. (2020). Model reference robust adaptive control of control element drive mechanism in a nuclear power plant. *Internat. J. Cont. Automat. Syst.* 18, 1651–1661. doi: 10.1007/s12555-019-0987-3

Perpetuini, D., Cardone, D., Filippini, C., Chiarelli, A. M., and Merla, A. (2021). A motion artifact correction procedure for fNIRS signals based on wavelet transform and infrared thermography video tracking. *Sensors* 21, 5117. doi: 10.3390/s21155117

Pham, P., Kim, G., Nguyen, Q., and Hong, K.-S. (2021). Control of a non-uniform flexible beam: identification of first two modes. *Internat. J. Cont. Automat. Syst.* 19, 3698–3707. doi: 10.1007/s12555-020-0913-8

Pinti, P., Aichelburg, C., Lind, F., Power, S., Swingler, E., Merla, A., et al. (2015). Using fiberless, wearable fNIRS to monitor brain activity in real-world cognitive tasks. *Jove-J. Vis. Exp.* 2015:e53336. doi: 10.3791/53336

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann. N Y Acad. Sci.* 1464, 5–29.

Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., et al. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *Neuroimage* 85, 64–71. doi: 10.1016/j.neuroimage.2013.06.062

Radhakrishnan, H., Vanduffel, W., Deng, H. P., Ekstrom, L., Boas, D. A., and Franceschini, M. A. (2009). Fast optical signal not detected in awake behaving monkeys. *NeuroImage* 45, 410–419. doi: 10.1016/j.neuroimage.2008.12.014

Rea, M., Rana, M., Lugato, N., Terekhin, P., Gizzi, L., Brotz, D., et al. (2014). Lower limb movement preparation in chronic stroke: a pilot study toward an fNIRS-BCI for gait rehabilitation. *Neurorehabilit. Neur. Rep.* 28, 564–575. doi: 10.1177/1545968313520410

Reyes, L. M. D., Bohache, K., Wijeakumar, S., and Spencer, J. P. (2018). Evaluating motion processing algorithms for use with functional near-infrared spectroscopy data from young children. *Neurophotonics* 5:025008. doi: 10.1117/1.NPh.5.2.025008

Robertson, F. C., Douglas, T. S., and Meintjes, E. M. (2010). Motion artifact removal for functional near infrared spectroscopy: a comparison of methods. *IEEE Transac. Biomed. Eng.* 57, 1377–1387. doi: 10.1109/TBME.2009.2038667

Safaie, J., Grebe, R., Moghaddam, H. A., and Wallois, F. (2013). Toward a fully integrated wireless wearable EEG-NIRS bimodal acquisition system. *J. Neur. Eng.* 10:056001. doi: 10.1088/1741-2560/10/5/056001

Santosa, H., Aarabi, A., Perlman, S. B., and Huppert, T. J. (2017). Characterization and correction of the false-discovery rates in resting state connectivity using functional near-infrared spectroscopy. *J. Biomed. Opt.* 22:055002. doi: 10.1117/1.Jbo.22.5.055002

Santosa, H., Hong, M. J., Kim, S. P., and Hong, K.-S. (2013). Noise reduction in functional near-infrared spectroscopy signals by independent component analysis. *Rev. Sci. Instrum.* 84:073106. doi: 10.1063/1.4812785

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., et al. (2014a). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27. doi: 10.1016/j.neuroimage.2013.05.004

Scholkmann, F., Metz, A. J., and Wolf, M. (2014b). Measuring tissue hemodynamics and oxygenation by continuous-wave functional near-infrared spectroscopy-how robust are the different calculation methods against movement artifacts? *Physiolog. Meas.* 35, 717–734. doi: 10.1088/0967-3334/35/4/717

Scholkmann, F., Spichtig, S., Muehlemann, T., and Wolf, M. (2010). How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiol. Meas.* 31, 649–662. doi: 10.1088/0967-3334/31/5/004

Schroeter, M. L., Zysset, S., Kruggel, F., and von Cramon, D. Y. (2003). Age dependency of the hemodynamic response as measured by functional near-infrared spectroscopy. *Neuroimage* 19, 555–564. doi: 10.1016/S1053-8119(03)00155-1

Seghouane, A., and Ferrari, D. (2019). Robust hemodynamic response function estimation from fNIRS signals. *IEEE Transac. Sign. Proc.* 67, 1838–1848. doi: 10.1109/Tsp.2019.2899289

Selb, J., Yucel, M. A., Phillip, D., Schytz, H. W., Iversen, H. K., Vangel, M., et al. (2015). Effect of motion artifacts and their correction on near-infrared spectroscopy oscillation data: a study in healthy subjects and stroke patients. *J. Biomed. Opt.* 20:056011. doi: 10.1117/1.Jbo.20.5.056011

Selesnick, I. W., Graber, H. L., Ding, Y., Zhang, T., and Barbour, R. L. (2014). Transient artifact reduction algorithm (TARA) based on sparse optimization. *IEEE Transac. Sign. Proc.* 62, 6596–6611. doi: 10.1109/Tsp.2014.2366716

Sherafati, A., Snyder, A. Z., Eggebrecht, A. T., Bergonzi, K. M., Burns-Yocum, T. M., Lugar, H. M., et al. (2020). Global motion detection and censoring in high-density diffuse optical tomography. *Hum. Brain Mapp.* 41, 4093–4112. doi: 10.1002/hbm.25111

Shukla, J., Barreda-Angeles, M., Oliver, J., and Puig, D. (2018). Efficient wavelet-based artifact removal for electrodermal activity in real-world applications. *Biomed. Sign. Proc. Contr.* 42, 45–52. doi: 10.1016/j.bspc.2018.01.009

Siddiquee, M. R., Hasan, S. M. S., Marquez, J. S., Ramon, R. N., and Bai, O. (2020). Accurate vigilance detection during gait by using movement artifact removal. *IEEE Access* 8, 51179–51188. doi: 10.1109/Access.2020.2980546

Siddiquee, M. R., Marquez, J. S., Atri, R., Ramon, R., Mayrand, R. P., and Bai, O. (2018). Movement artefact removal from NIRS signal using multi-channel IMU data. *Biomed. Eng. Online* 17:120. doi: 10.1186/s12938-018-0554-9

Sun, Y. B., and Zhao, Y. G. (2021). Distributed cubature kalman filter with performance comparison for large-scale power systems. *Internat. J. Cont. Automat. Syst.* 19, 1319–1327. doi: 10.1007/s12555-019-1054-9

Sutoko, S., Monden, Y., Funane, T., Tokuda, T., Katura, T., Sato, H., et al. (2018). Adaptive algorithm utilizing acceptance rate for eliminating noisy epochs in block-design functional near-infrared spectroscopy data: application to study in attention deficit/hyperactivity disorder children. *Neurophotonics* 5:045001. doi: 10.1117/1.NPh.5.4.045001

Sweeney, K. T., McLoone, S. F., and Ward, T. E. (2013). The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique. *IEEE Transac. Biomed. Eng.* 60, 97–105. doi: 10.1109/Tbme.2012.2225427

Tak, S., and Ye, J. C. (2014). Statistical analysis of fNIRS data: A comprehensive review. *Neuroimage* 85, 72–91. doi: 10.1016/j.neuroimage.2013.06.016

Takahashi, K., Tanaka, T., Nara, H., Kaneko, S., Inoue, M., Shimizu, S., et al. (2013). A modeling of cerebral blood flow changes due to head motion for fNIRS. *Internat. J. Optomechatr.* 7, 83–95. doi: 10.1080/15599612.2013.777493

Tang, M., Chen, Z., and Yin, F. (2020). Robot tracking in SLAM with Masreliez-Martin unscented Kalman filter. *Internat. J. Contr. Automat. Syst.* 18, 2315–2325. doi: 10.1007/s12555-019-0669-1

Vinette, S. A., Dunn, J. F., Slone, E., and Federico, P. (2015). Artifact reduction in long-term monitoring of cerebral hemodynamics using near-infrared spectroscopy. *Neurophotonics* 2:025004. doi: 10.1117/1.NPh.2.2.025004

Virtanen, J., Noponen, T., Kotilahti, K., Virtanen, J., and Ilmoniemi, R. J. (2011). Accelerometer-based method for correcting signal baseline changes caused by motion artifacts in medical near-infrared spectroscopy. *J. Biomed. Opt.* 16:087005. doi: 10.1117/1.3606576

Vitorio, R., Stuart, S., Rochester, L., Alcock, L., and Pantall, A. (2017). fNIRS response during walking - artefact or cortical activity? a systematic review. *Neurosci. Biobehav. Rev.* 83, 160–172. doi: 10.1016/j.neubiorev.2017.10.002

von Luhmann, A., Boukouvalas, Z., Muller, K. R., and Adali, T. (2019). A new blind source separation framework for signal analysis and artifact rejection in functional near-infrared spectroscopy. *Neuroimage* 200, 72–88. doi: 10.1016/j.neuroimage.2019.06.021

Wan, E. A., and Nelson, A. T. (2001). "Dual extended Kalman filter methods," in *Kalman Filtering and Neural Networks*, ed. S. HAYKIN (Hoboken, NJ: John Wiley & Sons, Inc).

Wei, Y., Wang, Y., Wang, G., and Yan, X. (2018). "Dual-threshold wavelet filtering based motion artifact removal for near-infrared spectroscopy long-term measurement," in *Proceedings of the BIBE 2018; International Conference on Biological Information and Biomedical Engineering, 6–8 June 2018*, Shanghai, 1–5.

Wilcox, T., Bortfeld, H., Woods, R., Wruck, E., and Boas, D. A. (2005). Using near-infrared spectroscopy to assess neural activation during object processing in infants. *J. Biomed. Optics* 10:011010. doi: 10.1117/1.1852551

Wyser, D., Lambercy, O., Scholkmann, F., Wolf, M., and Gassert, R. (2017). Wearable and modular functional near-infrared spectroscopy instrument with multidistance measurements at four wavelengths. *Neurophotonics* 4:041413. doi: 10.1117/1.NPh.4.4.041413

Yamada, T., Umeyama, S., and Ohashi, M. (2015). Removal of motion artifacts originating from optode fluctuations during functional near-infrared spectroscopy measurements. *Biomed. Optics Exp.* 6, 4632–4649. doi: 10.1364/BOE.6.004632

Yang, B., Jia, X. C., and Yang, F. W. (2021). Variational bayesian adaptive unscented kalman filter for RSSI-based indoor localization. *Internat. J. Contr. Automat. Syst.* 19, 1183–1193. doi: 10.1007/s12555-019-0973-9

Yucel, M. A., Selb, J., Cooper, R. J., and Boas, D. A. (2014a). Targeted principle component analysis: a new motion artifact correction approach for near-infrared spectroscopy. *J. Innov. Opt. Health Sci.* 7:1350066. doi: 10.1142/S1793545813500661

Yucel, M. A., Selb, J., Boas, D. A., Cash, S. S., and Cooper, R. J. (2014b). Reducing motion artifacts for long-term clinical NIRS monitoring using collodion-fixed prism-based optical fibers. *Neuroimage* 85, 192–201. doi: 10.1016/j.neuroimage.2013.06.054

Zafar, A., and Hong, K.-S. (2017). Detection and classification of three-class initial dips from prefrontal cortex. *Biomed. Opt. Exp.* 8, 367–383. doi: 10.1364/BOE.8.000367

Zhang, Y. H., Brooks, D. H., Franceschini, M. A., and Boas, D. A. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *J. Biomed. Opt.* 10:011014. doi: 10.1117/1.1852552

Zhao, H. B., Brigadoi, S., Chitnis, D., De Vita, E., Castellaro, M., Powell, S., et al. (2020). A wide field-of-view, modular, high-density diffuse optical tomography system for minimally constrained three-dimensional functional neuroimaging. *Biomed. Opt. Exp.* 11, 4110–4129. doi: 10.1364/BOE.394914

Zhao, J., Qiao, J. R. M. T., Ding, X. T., and Liang, X. M. (2020). fNIRS signal motion correction algorithm based on mathematical morphology and median filter. *ACTA Optica Sinica* 40:2230002. doi: 10.3788/Aos202040.2230002

Zhao, H. B., and Cooper, R. J. (2018). Review of recent progress toward a fiberless, whole-scalp diffuse optical tomography system. *Neurophotonics* 5:011012. doi: 10.1117/1.NPh.5.1.011012

Zhao, H. B., Frijia, E. M., Rosas, E. V., Collins-Jones, L., Smith, G., Nixon-Hill, R., et al. (2021). Design and validation of a mechanically flexible and ultra-lightweight high-density diffuse optical tomography system for functional neuroimaging of newborns. *Neurophotonics* 8:015011. doi: 10.1117/1.NPh.8.1.015011

Zhao, Y., Sun, P. P., Tan, F. L., Hou, X., and Zhu, C. Z. (2021). NIRS-ICA: A Matlab toolbox for independent component analysis applied in fNIRS studies. *Front. Neuroinform.* 15:683735. doi: 10.3389/fninf.2021.683735

Zhou, L. F., Chen, C., Liu, Z. H., Hu, Y. Y., Chen, M. Y., Li, Y. F., et al. (2021). A coarse/fine dual-stage motion artifacts removal algorithm for wearable NIRS systems. *IEEE Sens. J.* 21, 13574–13583. doi: 10.1109/Jsen.2021.3069553

Zhou, Z. B., Wu, J., Li, Y., Fu, C., and Fourati, H. (2017). Critical issues on kalman filter with colored and correlated system noises. *Asian J. Cont.* 19, 1905–1919. doi: 10.1002/asjc.1545

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership